

A pudding worth the eating

The results of the World Conference on Science, which ended last week, should not be exaggerated. But they are a firm basis on which governments in the developing world can plan their future support for science.

It is easy — perhaps too easy — to be cynical about meetings such as the World Conference on Science, which ended in Budapest last week (see pages 100 and 101). A six-day talking shop that produces two fat documents of 'principles' and 'guidelines', but not a single clear-cut decision, nor any strategic plan with identified objectives, is guaranteed to raise eyebrows. And those who argue that a critical determinant of a country's scientific potential is the size of its research budget will inevitably be frustrated by a meeting that refuses even to recommend extra spending on science.

It is equally easy to be complacent. With delegates from more than 150 countries endorsing, after lengthy negotiations, the documents in question — the *Declaration on Science* and the *Framework for Action* — it is tempting to give both an authority that would be deceptive. For all the talk of a 'new social contract' between science and society embodied in these documents, so much is missing, in terms of the economic and political realities needed to put the principles they contain into practice, that they invite dismissal as little more than a raft of good intentions.

Both sets of criticisms are unfair. The answer to those who lament the lack of any clear agenda emerging from Budapest is that the meeting, as jointly conceived by the United Nations Educational, Scientific and Cultural Organization (Unesco) and the International Council for Science (ICSU), was never intended to achieve this. Rather, it was aimed at — and largely succeeded in — establishing a consensus on the essential components of policies intended to promote the scientific capabilities of any state, developed or (in particular) developing. This message was intended for policy-makers, not scientists. If there is one regret, it is that the senior politicians who attended were science and education ministers, not the finance ministers — or even heads of state — who hold the purse strings.

But the danger of overstating the significance of the outcome lies in ignoring the fact that the conference will be judged, not by the seductiveness of carefully crafted paragraphs, but by the practical changes they bring about. Many of the organizers rightly criticize the lack of impact of the last such conference, held in Vienna 20 years ago.

If their efforts are to avoid the same fate, as much effort needs to go into implementing the high-minded phrases expressed in Budapest as went into polishing them in the first place.

There is plenty of scope. The meeting highlighted many areas where constructive efforts are needed, from enhancing the position of women in science, through boosting collaborative efforts to provide research training, to exploring new, cost-effective funding mechanisms. It proposed concrete measures for doing this, from an international women's network, to funding science through debt relief. And it also suggested where responsibility for achieving some of these goals should lie; monitoring the implementation of these suggestions, providing it is done in a fully transparent manner, is now essential to ensure that those identified in this way live up to their responsibilities.

There were other positive signs from Budapest. Those who lamented the lack of ringing endorsements of new projects ignore the amount of work that took place in the corridors intended to ensure that such initiatives — ranging from a proposed synchrotron radiation facility in the Middle East to the possible setting up of an International Centre for the Communication of Science — see the light of day. Earlier preparations for the conference have already stimulated a constructive, and continuing, debate on the possibilities for regional collaboration. Perhaps most significantly, a number of countries and regions, such as India (see page 95) and Africa (see page 101), have already promised to use the consensus on principles and guidelines emerging from Budapest to catalyse their own efforts to develop a sustainable science base.

It is essential that this momentum be maintained. Part of the Vienna conference's failure was the way that responsibility for follow-up was left to United Nations agencies for whom the health of global science was never a top priority. This time round, even though aid agencies and organizations such as Unesco and ICSU can act as useful catalysts, responsibility for action must remain firmly with national governments and regional bodies. This is one pudding whose proof really will lie in the eating. □

Learning by incentive

University reforms would help Germany to combat its worrying shortage of bioinformaticists.

Germany is not isolated in its need to address a shortage of bioinformaticists to support increasing efforts in genomic and post-genomic research (see page 102). But neighbouring countries have reacted with somewhat greater agility to the scientific opportunities being opened up by the data already flooding in from the nearly completed human genome project. Switzerland, for example, has just opened its Swiss Institute for Bioinformatics (SIB) in Lausanne, an initiative supported by local universities and research institutes.

The SIB's activities include teaching undergraduate and post-graduate university courses, leading to a nationally recognized certificate. This is exactly the sort of national initiative that Germany needs to support the programme launched by its university granting

agency, the Deutsche Forschungsgemeinschaft. As the German research ministry is currently in the middle of developing a new genome research strategy, this is a good time for it to try some creative thinking along the lines of the Swiss model.

Good things can happen fast in Germany if the incentives are right; a few years ago, the research ministry's Bioregio programme, a sort of formal ranking of highly competitive regions, turned around the fortunes of German biotechnology within a year. But the attack must be on several fronts. Federal and *Länder* governments must pursue reforms to allow universities to offer competitive salaries and more flexible employment conditions. Given the worldwide shortage of those qualified to teach bioinformatics, it is important that candidates see a good reason to do so in Germany. □



US plans large funding boost to support nanotechnology boom

[WASHINGTON] The US government plans to launch a major, inter-agency nanotechnology initiative to nurture what officials describe as explosive growth in scientific interest in the behaviour of materials at the nanometre scale.

The National Science Foundation (NSF), which supports most university research in nanoscience and is likely to lead the initiative, reports that it can fund only 13 per cent of the grant applications it receives in the field, compared with the 40 per cent success rate in many disciplines at the agency.

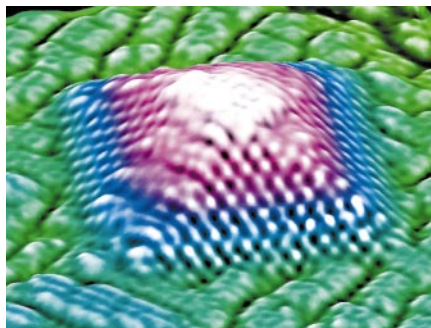
Competition for funds in the field is "absolutely ferocious," says Stan Williams, head of basic research at Hewlett-Packard and a keen supporter of the initiative. Williams prefers the term nanoscience to nanotechnology because the latter term has become tarred by fanciful claims on its behalf.

"Part of the problem is that nanotechnology has been over-hyped," he says, noting that "we have to fight against the distaste" that some feel, associating the term with visions of a factory housed in a matchbox.

Instead, a growing understanding of materials at the nanometre scale, where dimensions are comparable to the lengths of individual molecules, is likely to transform large-scale products and processes. "This doesn't need to be about small things, it could be about parts in a car," explains Mike Roco of the NSF's engineering directorate, who chairs an inter-agency working group that is planning the initiative. "The utilization of nanotechnology is very broad, but all fields of it use the same tools and methods."

Perhaps the most commercially spectacular application so far is the use of giant magnetoresistance — discovered in 1988 — in the reading heads of most computer disk drives. The related phenomenon of tunnelling magnetoresistance will shortly allow the production of fast and compact random-access memories for computers.

But, as Williams points out, imminent applications of nanotechnology extend far beyond the computer industry. Kodak, he says, is developing nanoscale particles called 'dygments' — a cross between powder pigments and molecular dyes — for use in printing images. Tyre manufacturers plan to mix nanoparticles of clay with tyre rubber, tying up loose ends of polymer molecules and greatly extending tyre life. And the four-fifths of possible drug therapies that can't be tested in patients because they are insoluble in water could be produced as nanoparticles sitting in a suspension in water, and could



Small can be profitable: a 'quantum dot' of germanium atoms, 10 nm across.

therefore become viable therapy candidates.

All of this potential is attracting attention in Washington, where support for a research initiative is growing. In the annual budget guidelines circulated to agencies in May, Jack Lew, the director of the Office of Management and Budget at the White House, and Neal Lane, director of its Office of Science and Technology Policy, identified nanotechnology as an area ripe for special inter-agency attention. Last month, hearings in both the House and the Senate outlined the special potential of the field.

All of this points to the inclusion of a major nanotechnology initiative in the budget proposal for the 2001 fiscal year, which President Bill Clinton presents to the Congress next February. A report soon to be

released by the inter-agency group calls for the initiative to double government spending on nanotechnology research from \$250 million to \$500 million over three years. Officials involved in planning the initiative hope that Clinton himself will announce it, possibly as early as September.

The NSF will spend \$80 million this year on nanotechnology research, while the Department of Defense spends \$60 million and the Department of Energy \$54 million. These figures suggest that total US government spending has doubled since 1997, when a study by the World Technology division of Loyola College estimated total government spending at \$116 million. Japan spent \$128 million in 1997 and Western Europe \$120 million, according to the study. But Rico says that spending in each region of the world has grown sharply since then.

"The European Union, Germany and Japan each have very focused efforts in nanotechnology," Williams says, adding that the US government research effort has lacked coordination, so that "some areas of research are being ignored completely".

However, the Loyola study found that the United States led the world in the synthesis and assembly of nanostructures, and in high-surface-area materials. It shared the lead with Europe in coatings and biological applications, while Japan led in 'nanodevices' and consolidated materials.

Colin Macilwain

India backs World Conference decisions

[NEW DELHI] The Indian government has announced that it intends to convene a meeting of the country's top scientists to draw up a national plan of action to implement recommendations endorsed by the World Conference on Science, which ended in Budapest last week.

The recommendations are included in two documents, *Declaration on Science and the Uses of Scientific Knowledge* and *Science Agenda: Framework for Action*. These were unanimously endorsed by delegates at the end of the six-day meeting, which was jointly organized by the United Nations Educational, Scientific and Cultural Organization (Unesco) and the International Council for Science (ICSU) (for full report, see pages 100 and 101).

The Indian action plan will be submitted to the Unesco general conference in October, according to Muril Manohar Joshi, India's minister for science, who led the country's delegation in Budapest.

At a press conference in New Delhi, Joshi said that India was pleased with the outcome of the conference, and especially with the widespread support it received for its proposals for protecting traditional knowledge systems, compensating developing nations for the loss they suffer through the brain drain, and the possible regular publication of a World Technology Report.

Given that India's proposals on these topics had been adopted, Joshi said that the country would make a concentrated effort to ensure that steps are taken, in consultation with other countries in the region, to put into practice the principles and guidelines contained in the two final documents.

"We will immediately start those activities for which external support is likely to become available," says Valangiman Ramamurthi, secretary to the Department of Science and Technology. **K. S. Jayaraman**

Protesters seek US ban on embryo stem-cell work...

[WASHINGTON] More than 100 anti-abortionists, led by a conservative US senator, last week released a statement arguing that research on human embryonic stem cells is scientifically unnecessary and urging Congress to fund alternatives.

The statement came as debate on the issue heated up in Washington in anticipation of the release this month of a report from the National Bioethics Advisory Commission. The commission is expected to recommend that the government should fund both the derivation of stem cells from embryos — which needs the embryos to be destroyed — and the use of the cells in research.

The National Institutes of Health (NIH) is also due soon to issue guidelines for the investigators it funds. The NIH has already said it will fund research on stem cells, but not their derivation from human embryos (see *Nature* 397, 185 & 399, 292; 1999).

The group that signed last week's statement included Frank Young, a former commissioner of the Food and Drug Administration (FDA), C. Everett Koop, a former surgeon general, and Edmund Pellegrino, director of the Center for Clinical Bioethics at Georgetown University, Washington DC.

The statement was coordinated by the Center for Bioethics and Human Dignity in Bannockburn, Illinois. It was released at a press conference at which Senator Sam Brownback (Republican, Kansas) said that research on human embryonic stem cells is

"immoral, illegal and unnecessary".

Brownback and the signatories to the statement contend that research on stem cells and their derivation from human embryos violates a federal law forbidding research in which embryos are destroyed or discarded. They said that recent research showing that stem cells isolated from adult tissues can give rise to differentiated cells removes the need for research on embryonic stem cells.

Young, who served as FDA commissioner under President Ronald Reagan, compared the development of cell and tissue therapies from embryonic stem cells to the making of saddles from human skins by Nazi Germans. He predicted "a backlash of the American people" if federal funding for the work proceeds. Fear of such a backlash has made corporate America "very apprehensive" about funding the research, he claimed.

But John Gearhart, a professor of obstetrics and gynaecology at Johns Hopkins University School of Medicine in Baltimore, says it is unproven that stem cells from adult tissue can generate all the tissues of the body.

"Clearly, embryonic stem cells produce more types of cells than the [stem] cells obtained from adult tissue," says Gearhart. "We should leave our options open" by pursuing embryonic stem-cell research. Gearhart announced last November that he had isolated stem cells using germline cells from aborted fetuses (see *Nature* 396, 104; 1998).

Meredith Wadman

... as Wisconsin researcher awaits go-ahead

The current debate on the ethics of stem-cell research (see above) was launched by the work of James Thomson, who isolated human embryonic stem cells from the inner cell mass of embryos left over at fertility clinics. Thomson relied on private money from Geron, a Californian biotechnology company, to achieve his feat. Private work with stem cells is legal in the United States, and is likely to remain so. The debate centres on whether such work should receive federal funding.

But Thomson, a developmental biologist at the University of Wisconsin-Madison, says his private-sector financing has

left him far from untouched by the ban on federal funding.

This is because the ban requires strict separation of stem-cell work from federally funded work. Thomson had to set up a physically separate laboratory in 1997 when he launched his effort to isolate the cells.

That laboratory, at the Clinical Science Center at the University of Wisconsin Hospital and Clinics in Madison, is not as well equipped as his main laboratory. Thomson has been forced to limit his current research to basic tissue-culture manipulation experiments. Tantalizing challenges, such as

identifying the genes that direct stem cells to become one tissue type or another, are being deferred through lack of equipment.

Thomson is therefore anxiously awaiting the publication of the NIH guidelines that will tell investigators how to proceed when applying for, and using, federal funding for the research. Despite the Congressional ban, the biomedical agency thinks it is permitted to fund research on stem cells, following advice from lawyers in the Department of Health and Human Services. The publication of guidelines will signal that the NIH is ready to begin issuing grants. **M.W.**

US universities request special help as visa quota is reached early

[WASHINGTON] Despite legislation that nearly doubled the number of computer programmers, academic researchers and other high-tech workers allowed into the United States in 1999, the annual quota of H-1B visas has been reached with three months still remaining in the fiscal year.

All 115,000 visas in the category were given out by the middle of last month, and the US Immigration and Naturalization Service (INS) says it will not grant any more until October.

Congress raised the H-1B quota from 65,000 last summer after a lengthy political fight that pitted US labour advocates against the computer software industry, which is keen to hire more foreign programmers (see *Nature* 394, 610; 1998).

American universities use only a small fraction of the H-1B allocations each year for faculty appointments. But evidence collected by the Washington-based Association of International Educators (NAFSA) suggests that universities bear a disproportionate amount of the pain when visas run out early. Because the academic fiscal year ends in July, new faculty tend to be hired just when the H-1Bs are becoming scarce.

Representatives from NAFSA, the College and University Personnel Association, and the Association of American Universities met with key legislators last month to plead their case. Among the proposals to address the special needs of higher education would be a plan to reserve a certain number of H-1B visas for academic use.

But this would have to be done through legislation, and observers say there is little appetite on Capitol Hill this year for another battle over high-tech visas.

Senator Phil Gramm (Republican, Texas) plans to introduce a bill this month calling for the H-1B cap to be raised permanently to 200,000. But this would almost certainly be opposed by the White House and many pro-labour Democrats, and is given little chance of passing.

Meanwhile, Lamar Smith (Republican, Texas), who chairs the House of Representatives subcommittee on immigration and was a key figure in last year's H-1B compromise, has been more concerned recently about the misuse of visas, particularly in India, which he says is taking a significant portion of the annual quota.

According to an investigation cited by Smith, 45 per cent of H-1B applications examined by a US consulate in India could not be authenticated, and 21 per cent were fraudulent.

Tony Reichhardt

EMBO backs single electronic repository

[PARIS] The European Molecular Biology Organization (EMBO) is supporting the launch of a global web repository for literature in the life sciences, in cooperation with the US National Institutes of Health (NIH).

The concept of a database that would be freely accessible to all — called 'E-Biomed' — has been instigated by Harold Varmus, the director of the NIH. He has proposed that it be taken forward via an international coalition (see *Nature* **398**, 735 & **399**, 8–9; 1999).

EMBO will "aim to be a partner in the start-up international governing body that will be formed by interested parties," says Frank Gannon, executive director of EMBO. "EMBO's position is that a single searchable location for all scientific data of relevance to life sciences is a very desirable goal."

Under the proposal, EMBO scientists would help to administer an assessment procedure for the repository. The computing infrastructure for the European arm would be handled by the European Molecular Biology Laboratory's outstation, the European Bioinformatics Institute, in Cambridge, UK.

EMBO's vision of the repository differs from that of 'E-Biomed', however. For example, it wants the repository not to be restricted to the biomedical literature, but to include all of the life sciences, including plant biology and biotechnology.

According to internal EMBO documents, the organization's main concern is that the 'E-Biomed' proposal could result in a "loss



Gannon: concern over quality control.

of quality control" that could seriously damage molecular biology.

The 'E-Biomed' proposal allows authors to submit to either an unrefereed e-print server or a peer-reviewed server — in effect, a large electronic journal. But EMBO rejects an unrefereed repository, and also questions the need for full-blown peer review within 'E-Biomed'.

An unrefereed site along the lines of the Los Alamos e-print server in physics will not work in biology, asserts Gannon. The interpretation of biological research is much more subjective — and controversial — than that of physics, he says, and is highly dependent on the controls carried out.

"I think that the material on an e-print server for biology which is unrefereed, unassessed or 'un-quality controlled', really has no more status than press releases," says Gannon.

Instead, EMBO proposes that all papers submitted to 'E-Biomed' should be "assessed" for acceptance by a mosaic of learned societies to ensure that they constitute reasonably sound science. In practice, scientists submitting to the repository could,

for example, send their paper to EMBO.

EMBO would send the paper to one of its 1,000 members who, according to Gannon, would ask: "Are there good reproducible methods? Are the data presented clearly? Are the conclusions reasonable, and not extrapolated to folly?" Papers meeting these criteria would receive an 'EMBO approved' stamp.

"There is a big difference between this and refereeing in the journal sense," he says. "A referee in a journal is also asking: is this sound? But they are integrating it into a scale of quality as judged by universal interest, interest for readers, and ranking compared with other material submitted."

Under the EMBO proposal, the repository would not introduce such a peer-review system. It argues that this would be unnecessarily complicated, and is best left to existing journals.

Gannon predicts that journals of low quality and circulation will disappear, but the top journals will remain. "The leading journals carry a message that this is work that has been upgraded/improved/approved by the refereeing system; they will be seen as a real asset in a way that has been taken for granted so far."

The need, says Gannon, is not to reinvent the journals system within 'E-Biomed', but to link in the journals, perhaps by licensing material from them. "We are very anxious to defend top journals and scientific societies; to lose that at a stroke would be bad for science."

Declan Butler

Commission to report on overhaul for French research structure

[PARIS] The creation of a 'parliament' of French scientists, to act as a bridge between research bodies and governments, is likely to be among the recommendations expected this month from a parliamentary commission on French research set up earlier this year by the prime minister, Lionel Jospin.

The parliamentary commission — headed by members of the National Assembly, Pierre Cohen and Jean-Yves Le Déaut — wound up at the end of last month with a national colloquium at the Sorbonne in Paris.

The outcome is expected to provide a basis for restarting discussions about reforms that have stalled for almost a year because of stiff opposition in the research community to proposals from the research minister, Claude Allègre.

The commission itself is not expected to make its recommendations to the government until later this month. As a result, the detailed status of the proposed 'parliament' of scientists remains unclear. But Le Déaut says it will assess national

research programmes and comment on the distribution of funding.

Allègre's proposals to make the universities the central plank in the French science system, in place of the national research agencies such as the Centre National de la Recherche Scientifique (CNRS), are unlikely to be backed by the commission.

Le Déaut argues that "the respective roles of the agencies and universities are well established: the agencies should steer nationally and the universities locally", but he supports greater cooperation between the two.

Several delegates at the Sorbonne colloquium argued that many universities lacked the capacity of the research agencies to organize national research strategies. Some called for a complete reform of the university system.

This is already essential, say some observers, because a shift in research effort from research agencies to universities is taking place, as recruitment to universities is almost three times that at the agencies.

Closing the meeting, Le Déaut said there was a need to end 'old boy' networks, reduce bureaucracy, and modernize what he described as the excessively top-down structure of evaluation and other commissions.

But he argued that these problems were not restricted to the universities, as they were also common within agencies. Indeed, the commission seems likely to back plans by Allègre for an overhaul of research evaluation, with greater input from scientists in other countries. Evaluation of laboratories and staff would be carried out separately, with the former shifting to a system based more on selection of research projects.

The commission seems unlikely to reject a proposal to create a single employment status of lecturer/researcher. At the moment, some agency scientists enjoy full-time research while their colleagues in the universities struggle under heavy teaching loads. It is also expected to recommend a reduction in teaching loads and greater opportunities for transfers between the universities and research agencies. **Eric Glover**

Job losses feared as UK marine science centre overspends

[LONDON] Britain's Natural Environment Research Council (NERC) has discovered that one of its research centres — the Centre for Coastal and Marine Sciences (CCMS) — has overspent its budget by £2.4 million (US\$3.77 million).

A spokesman for the staff labour union, the Institute of Professionals, Managers and Specialists, says representatives from the staff and management will this week attempt to reduce the budget without making redundancies at the centre's three laboratories in Cheshire, Oban and Plymouth.

But observers doubt whether redundancies can be avoided. Some estimates suggest that up to 30 jobs may have to go to compensate for the overspend, which amounts to around 7 per cent of the centre's annual budget. The CCMS has already slimmed down considerably as part of a restructuring exercise one year ago.

It is unclear how or why the overspend occurred, or why the shortfall was not identified earlier. Some staff blame defective accounting software. But most agree that the centre's increasing reliance on income from short term contracts is an important factor — and needs to be reviewed.

One CCMS scientist says he believes the overspend occurred partly because of "optimistic projections" of income from contract research, which accounts for around 40 per cent of the centre's income. The rest of the income is derived from the NERC as core research funding.

This view is shared by the NERC's chief executive, Sir John Krebs. In a statement last week, Krebs said that "all NERC's centres and surveys depend to a greater or lesser extent on external funding... as a result we expect fluctuations in the level of funding, which can sometimes cause cash flow difficulties".

He added: "This is exacerbated by the move away from large, longer term strategic contracts from government departments to smaller, shorter term contracts to address specific issues. This is an issue for the science base as a whole."

But one senior NERC-funded scientist claims that the research council's policies are also partly to blame for the funding shortfall — particularly the decision to allow core funding for marine sciences to decline for many years.

In addition, the CCMS's three laboratories are all struggling to find contract research for which they have to compete with universities.

Ehsan Masood



White out: widespread coral bleaching would pose a severe threat to ecosystems and local economies.

Global warming 'could kill most coral reefs by 2100'

[SYDNEY] An Australian scientist has identified global warming as the most likely culprit for last year's widespread coral bleaching, and predicts that similar events are likely to occur annually in most tropical oceans within 30–50 years.

The warning came this week from Ove Hoegh-Guldberg of the University of Sydney, who has studied for the past 15 years how the normally brilliant colours of coral turn white. He predicts that coral reefs "could be eliminated from most areas by 2100".

The potential impact on economic activity on reefs, especially fishing and tourism, is substantial. There are also implications for policies to curb global warming. The prediction is likely to have particular impact in Australia, where the government remains sceptical of a link between increasing temperatures and environmental degradation.

Coral obtain foods through the algae that live symbiotically within them. Bleaching occurs when the algae are expelled owing to damage by light at higher than normal temperatures, leaving stark, white skeletons.

Hoegh-Guldberg and Sandra Ward of the University of Sydney, and Peter Harrison of Southern Cross University, obtained results from six tanks placed on Australia's Great Barrier Reef. Corals were studied as the temperature of the seawater was artificially increased.

Once above the ambient 26–28 °C, there was a 10 per cent decrease in the rate of fertilization. At 32 °C, the rate of reproduction dropped dramatically to 40 per cent, and at 34 °C there was almost none.

Other evidence linking warming with bleaching came from satellite measurements of the temperature of the sea surface, gathered by the US National Oceanographic and Atmospheric Administration. During last year, the worst on record for bleaching, wher-

ever the temperature was only one degree above ambient, mass death of corals occurred.

Now, in what is claimed to be the first application of computer models to coral reef research, Hoegh-Guldberg has projected how the climate will change in regions where corals grow. He claims to have shown that, unless global warming is arrested, coral bleaching will occur more frequently, and more intensely, until by 2030 it will appear every year.

Every coral reef examined showed the same drastic trend, with consistency between the major oceans, although the rate of bleaching onset differs. Caribbean and Southeast Asian reefs would be hit first with annual bleaching by 2020, whereas central Pacific reefs would not be affected for another two decades, it is predicted.

The Great Barrier Reef sits between the two extremes, with annual bleaching being predicted by 2030.

"The rapidity and extent of the changes, if realized, spell catastrophe for tropical marine ecosystems everywhere, and suggest that unrestrained warming cannot occur without the complete loss of coral reefs on a global scale," says Hoegh-Guldberg.

The study was accepted last week for publication in the Australian journal, *Marine and Freshwater Research*. The research was financed by the environmental organization Greenpeace.

Some remain sceptical about the predicted frequency of coral bleaching events on the Great Barrier Reef. "To conclude that coral bleaching is due to global warming is very speculative and highly uncertain at this stage," says Malcolm McCulloch, an Earth scientist at the Australian National University.

But Terry Done, of the Australian Institute of Marine Research, describes the report as "quite credible".

Peter Pockley

NIH office plans research on misconduct

[SAN DIEGO] The Office of Research Integrity (ORI) of the US National Institutes of Health (NIH) is planning to launch a research programme to conduct studies on scientific misconduct and integrity issues. The programme is being called an 'invisible college', with scientists from various disciplines at different institutions independently studying misconduct.

About 15 representatives of organizations such as the National Science Foundation (NSF), the Office of the Inspector General, the NIH, the Association of American Medical Colleges and the Federation of American Societies for Experimental Biology (FASEB) attended the first planning meeting last month in Maryland.

Mary Scheetz, an information scientist at ORI's division of policy and education who helped organize the meeting, said there was strong support for scholarly studies on the topics. "There is a lack of solid research on scientific misconduct and research integrity," says Scheetz. "In order for agencies to make effective policies, they need to make decisions on solid facts. The idea is to produce a research programme to produce that data."

ORI is the lead agency addressing research misconduct occurring under NIH grants, but relies on institutions funded by the NIH to have policies and procedures to investigate allegations of fabrication, falsification or plagiarism.

When universities find misconduct, ORI conducts its own investigation and issues sanctions, including debarment from receiving federal funds. The NSF has similar procedures, but relies on its Office of the Inspector General to investigate.

Everyone at the meeting "agreed this is something that is important," says Rachelle Hollander, director of the NSF's program on societal dimensions of engineering, science and technology.

The federal watchdogs involved in the planning effort are keen to move beyond the image of being 'fraud police'. "We want to work with the extramural community and develop an agenda for research on misconduct," says Chris Pascal, acting director of ORI.

"We have passed the point of anecdotal evidence," added Scheetz. But she said the topics are "slippery" to study, because of the

nature of the subject and the potential negative impacts for the careers of those involved. As one scientist noted: "We don't reward whistleblowers; we destroy them."

Ruth L. Fischbach, senior adviser for biomedical ethics for the NIH's director of extramural research, said she is encouraged by the idea of research on misconduct. "This is an opportunity to look at the roots and learn to be proactive to try to prevent misconduct from occurring," said Fischbach. "I have found there is a lot of secrecy within institutions about revealing the level of misconduct."

ORI is already working with a private research institution to examine academic medical centre guidelines for research staff, including authorship, retention and recording of data, and intellectual property rights. This study may lead to a scientific conference on associated issues.

At its next meeting on the misconduct research programme in November, officials hope to commission research projects before a national conference in late 2000. The studies will be paid for initially with agency discretionary funds, officials say. **Rex Dalton**

NASA axes comet mission as delays to other projects prove costly

[WASHINGTON] Faced with cost overruns and expensive delays to other, higher-priority science projects, the US space agency NASA has cancelled a \$240 million comet landing mission that had been proposed for early in the next decade.

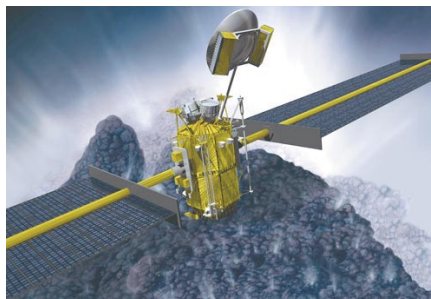
The Champollion spacecraft was to have been launched in 2003 as part of NASA's 'New Millennium' programme to demonstrate advanced spacecraft technologies.

It would have met up with Comet Tempel 1, anchored itself to the comet's nucleus and drilled for material below the surface to validate devices for use on future missions.

But Champollion was not primarily a science project and was still early in the planning stages — less than \$10 million has been spent on preliminary studies — so it became vulnerable when NASA's science office ran into financial difficulties.

Normally, the agency's \$2.1 billion science programme would be flexible enough to absorb tens of millions of dollars in unexpected costs in any given year, says Edward Weiler, head of NASA's space science office. But delays to the \$1.5 billion Chandra X-ray observatory, resulting partly from problems with the upper-stage rocket, will add \$100 million over the next few years.

Splitting a repair of the Hubble Space Telescope into two shuttle missions, partly because of earlier slips in the schedule, added \$26 million. The \$500 million Gravity



Grounded: Champollion was to have drilled under the surface of Comet Tempel 1 to collect material.

Probe-B satellite, which was scheduled to reach orbit next year, is \$20 million over budget, and NASA's Mars exploration programme was becoming low on reserves. Taken together, says Weiler, these were "not the usual year's money problems".

Champollion design work conducted at the Jet Propulsion Laboratory in California can still be applied to other missions, including an expedition to Jupiter's moon Europa. "I'm sure NASA will get around to doing a major comet mission sometime in the future," says Weiler.

For now, though, Europe's Rosetta mission, planned for launch in 2003, has that territory to itself. Champollion began life as the US contribution to Rosetta, but NASA backed out as a major partner three years ago to pursue its own comet lander as

a low-cost New Millennium spacecraft (see *Nature* 383, 469; 1996).

US scientists will still be joint investigators on Rosetta, which will land a probe on the surface of Comet Wirtanen in 2012. Weiler says the existence of the international mission was another factor against keeping Champollion alive.

The Planetary Society, a California-based space advocacy group, was not happy with NASA's decision, and called on the US House of Representatives Science Committee to help restore money for Champollion.

In a statement, the society's director Louis Friedman said that penalizing a successful science and technology mission to pay for other programmes that are experiencing cost overruns "sets a disturbing precedent".

But NASA could not count on a Congressional rescue in a year when lawmakers are warning of a difficult appropriations battle to meet spending caps. Weiler says that cancelling Champollion "goes a long way" towards solving his impending budget problem, if Chandra launches on 20 July as planned.

The demise of the comet lander is a lesson in the difficulties of managing many "better, cheaper, faster" space missions at once. As one congressional staffer says: "There isn't the play in the NASA budget that there had been in the past." **Tony Reichhardt**



The full texts of the two final statements endorsed by the World Conference on Science can be found on *Nature's* WCS website at <http://helix.nature.com/wcs>. Further information about the conference can also be found on <http://www.unesco.org/opi/science>. The following reports have been written by David Dickson, Natasha Loder and Ehsan Masood.

Guidelines agreed for new social contract

The World Conference on Science closed in Budapest last Thursday (1 July) with 1,800 delegates from more than 150 countries agreeing on principles and guidelines for implementing what the organizers described as 'new social contract' between science and society.

Included in the specific suggestions were that countries should provide increased support for the networking of graduate and postgraduate institutions, that the training of scientific journalists and communicators should be enhanced, and that "adequate participatory mechanisms should be instituted to facilitate democratic debate on scientific policy issues".

One point of agreement was that 'ethics and social responsibility' should be an integral part of the education of all scientists. Existing panels in Unesco and the International Council for Science (ICSU), the co-organizers of the conference, have been given responsibility for following up on this issue.

Another proposal was that campaigns should be launched at national, regional and global levels to raise awareness of the contributions of women to science and technology, and that an international network of women scientists should be set up.

Both of these ideas figure among a set of recommendations intended to enhance the position of women in science that were

adopted after vigorous lobbying by women's groups (see opposite).

The guidelines also encourage "special efforts" to ensure the full participation of disadvantaged groups — implicitly including the disabled — in science and technology. Such groups would be represented in policy-making bodies and forums.

The conference organizers had taken care to avoid explicitly committing participants to calls for any increased funding, not only from the industrialized to the developing nations, but also by developing nations themselves.

There was resistance to calls from some nations, in particular delegates from some Muslim states, that all countries should aim eventually to spend one per cent of their gross national product on research and development, a target that had been agreed at the last global science meeting, held 20 years ago in Vienna (see *Nature* **400**, 8; 1999).

However, the final conference documents did agree in general that "innovative and cost-effective mechanisms" for funding science should be examined for implementation by "relevant institutions at the regional and international levels".

The conclusions of the six-day meeting were expressed in two documents, *Declaration on Science and the Use of Scientific Knowledge*, and *Science Agenda: Framework*

for Action, known as the *Declaration* and *Framework* respectively.

Both were adopted unanimously by the final plenary session after several months of consultation with the member states of Unesco and other organizations, and a hectic two days of final drafting.

The first is a general statement of principles about the importance of science, as well as the need to respect a new 'contract' in which society pledges to continue to support science, while in return scientists agree to accept and respect their responsibilities.

The *Declaration* urges "the nations and the scientists of the world" to "acknowledge the urgency of using knowledge from all fields of science in a responsible manner to address human needs and aspirations without misusing this knowledge". It also says that helping to create a critical mass of national research in the sciences through regional and international cooperation is especially important for small states and least developed countries.

The second document is intended to provide guidelines through which the principles in the first document can be implemented by national governments, international organizations, professional scientific bodies — indeed all those keen to promote a responsible relationship between science and society.

Full text: <http://helix.nature.com/wcs/02-1g.html>

Conference a 'success' despite lack of support for global fund

One way of establishing whether the World Conference on Science was a worthwhile event is to ask delegates whether they would go to another such gathering. In a straw poll conducted shortly after the end of the meeting, all of those questioned — from both rich and poor countries — replied without hesitation that they would.

Indeed, Lieutenant General Muhammad Noor Uddin Khan, minister for science in Bangladesh, said that such meetings should be held more often "and not every 20 years". Even members of the US delegation, often among the most sceptical participants, agreed that the conference had been worthwhile.

US officials singled out for particular praise the fact that an international meeting of this kind had been organized jointly with a non-governmental organization, the International Council for Science (ICSU). "This has not happened before," said one official. "It shows that civil society can

become part of the United Nations system. It's a new way of doing business. It could be a model for future meetings."

For delegates from the developing countries, the conference provided the rare opportunity of direct access to key policy-makers and heads of funding agencies from developed countries, as well as some insight into the unfamiliar world of international diplomacy. At the same time, however, there were some inevitable disappointments.

The delegation from Bangladesh, for example, in common with those from many developing countries, returned home disappointed at the lack of agreement on a new global fund for science. Many delegates from poor countries had come to Budapest genuinely believing that such a fund might be agreed upon.

Some delegations from developing countries also voiced criticism of what they considered to be inadequate preparation for the conference. They expressed concern at

the absence of official preparatory meetings, usually an important feature of UN conferences.

Such meetings provide countries with a formal mechanism to comment on draft documents. Perhaps more importantly, they are occasions on which groups of countries can forge alliances around common demands. Indeed, it is often at preparatory meetings that the seeds of potential conflict with the developed world are first sown.

But the conference itself passed off relatively cordially. One reason was the absence of a strong developing country alliance. Perhaps another reason, according to a delegate from Senegal, was that many developing country delegates — who were drawn mainly from science ministries — had never been to an international meeting of this type, and therefore lacked the requisite diplomatic lobbying and language skills.

Full text: <http://helix.nature.com/wcs/1news/02-1b.html>

Women celebrate successful campaign on gender issues

Women's groups attending the World Conference on Science were opening the champagne on Wednesday night (30 June) as news began to leak to them from the committee drafting the final conference documents that they had been successful in significantly enhancing all references to the need to improve the position of women in science.

There had earlier been warnings that women and women's issues might be marginalized. But after a determined campaign and the final adoption of the *Declaration and Framework for Action* (see opposite), female participants in the conference were unanimous in their enthusiastic approval of the final version of the two documents.

"It is the first time that [the gender issue] has entered the world science agenda," says Sjamsiah Achmad, of the Indonesian Institute of Technology, in Jakarta, who had chaired the conference's 'thematic session' on gender-related issues. "We managed to get a whole new paragraph inserted into the *Framework for Action*. It is extraordinary."

Wati Hermawati, another member of the Indonesian delegation, says she will produce gender indicators, through her work for the National Focal Point for Gender Science and Technology at the same institute. Statistics and data will be collected in areas such as the education and careers of women in science, and the impact of science and technology on women.

Amalia Bosia, a molecular biologist from the University of Turin and a member of the Italian delegation, says that the final documents adopted by the meeting appear to have taken into account some of the views expressed to Unesco and ICSU during the preparatory process. These included declarations from seven regional forums that had looked at the issue of women in science. "Some of the things that went in [to the documents] may be obvious, but the important thing is that they came from a general opinion," says Bosia.

The success of the women's lobby at the conference was the result of a campaign that started more than a year ago, when representatives of the UN Development Fund for Women (UNIFEM) and the Once and Again Future Action Network (OFAN) decided that the conference was important.

These groups formed a gender networking group at the conference. The group lobbied non-governmental organizations, placed representatives in thematic meetings, analysed lists of delegates for gender bias, and even noted down the number of times that gender was mentioned during the plenary sessions.

Full text: <http://helix.nature.com/wcs/1news/02-1d.html>

Declaration the 'backbone' for cooperation in Africa

Research ministers from 50 African states are planning to meet in Cairo next January under the auspices of the Organization of African Unity (OAU) to discuss drawing up a protocol for scientific collaboration across the continent.

Such a protocol, which it is hoped will be signed by heads of state, could represent one of the more concrete follow-up activities to make use of the guidelines for promoting science in the interests of development agreed at the World Conference on Science.

The ideas enshrined in the declaration agreed at the Budapest meeting are likely to provide the "backbone" to such an agreement, says Henri Hogbe Nlend, minister for science and technology in Cameroon. Nlend, who will chair the Cairo meeting, describes it as an "African projection" of the conclusions of the world conference.

Nlend describes the conference as having been successful, partly because many of the ideas that African delegations had been keen to promote — such as that of science being "part of the common heritage of mankind" — were recognized in the conclusions.

Like others, he would have liked to have seen additional points included in the final declarations, such as explicit support for the need to increase funding for science. He points out that the last such meeting, in Vienna in 1979, endorsed a target of spending one per cent of gross national product on research.

"A figure like that creates a target by which we can measure our efforts," he says. "You can then tell governments that the money they allocate to science is too small in comparison to what other countries are spending."



Research collaboration: African protocol would put conference principles into practice.

But Nlend says he is pleased that the final declaration acknowledges the need for new funding mechanisms at national and regional levels. "There is a strong appeal to increase money for science and technology" he says. "Any new mechanism that is launched does not have to be a global fund."

Nlend says he is also pleased that the responsibility for follow-up activities has been broadened, allowing regional organizations (such as the OAU) to take up the challenge. He is less happy with those parts of the declaration that deal with intellectual property rights. He feels that the declaration skirts around issues raised by the privatization of research, for example in health provision.

"Such fields are more and more controlled by multinational corporations," says Nlend. "Delegates have spoken about the dangers of this, and the issue has been referred to in the final documents, but the wording is not strong enough."

Full text: <http://helix.nature.com/wcs/1news/02-1c.html>

US proposes focus on science teaching

The United States, currently considering the terms of a potential bid to rejoin Unesco, has suggested that the organization should consider "resurrecting" a previously successful division of science education within its education sector.

The suggestion came in a statement to the plenary session of the World Conference on Science by Michael Southwick, US deputy assistant secretary of state, who described the Budapest meeting as "Unesco at its best".

The United States withdrew from the organization in the mid-1980s over complaints of maladministration and over-politicization. A proposal last year from the State Department to ask Congress for funds to rejoin was shelved after discussions with

the Office of Management and Budget concluded that the timing was premature.

Southwick pointed out that there had been a time when the education sector had had an "extraordinarily active and effective" division of science education. "Might it be time to resurrect it?"

His remarks reflected strong feelings about the need to boost science education worldwide held by members of the US delegation, including Bruce Alberts, president of the National Academy of Sciences, and Leon Lederman, former director of the Fermi National Laboratory.

Lederman said scientists needed to communicate the importance of nurturing scientific literacy in schools.

Full text: <http://helix.nature.com/wcs/1news/02-1i.html>

German agency to boost bioinformatics

[MUNICH] Only weeks after the publication of a paper pointing to major deficits in Germany's genome research effort, the Deutsche Forschungsgemeinschaft (DFG), Germany's main funding agency for university research, has launched its own programme in bioinformatics.

The DFG is keen to encourage German universities to shed rigid, discipline-orientated structures that have prevented them from embracing bioinformatics as an important new academic discipline, and done little to lessen the chronic shortage of German bioinformaticists.

The speed of the DFG's decision, and the flexible nature of the programme itself, have both surprised and pleased German scientists. But the initiative is also a deliberate test of the DFG's ability to practise what it preaches and display the flexibility required to run a loosely structured programme.

The programme will consider for funding any proposal that could help to increase German competitiveness in bioinformatics. Two or three of the best proposals will share DM50 million (US\$26 million) over five years; the money may be spent on either research infrastructure or salaries.

Universities are being called upon to link

up with local non-university research institutes to create regional projects for either establishing or building up permanent research and training centres in bioinformatics. Such projects, says the DFG, may "even go as far as cooperation with industry".

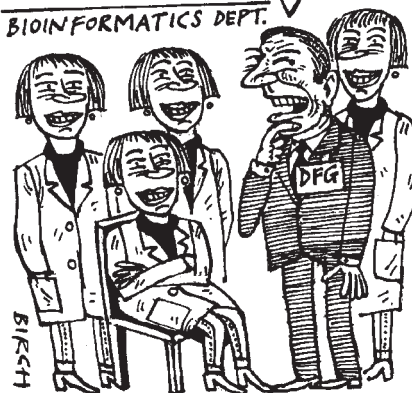
Winning proposals will be those that can persuade the DFG that they will create internationally competitive bioinformatics centres which will ultimately survive on funds from other sources.

"Winners will have to demonstrate convincingly that their proposal will result in structural changes" within universities, says the DFG's Andreas Engelke. Their aim must be "to provide a complete environment for bioinformatics, from undergraduate and postgraduate teaching to research programmes".

The DFG is not used to handling such a loosely defined programme. But its launch is partly a response to criticism from an independent evaluation committee, which concluded that the DFG's rigidity was an obstacle to supporting new interdisciplinary subjects (see *Nature* **399**, 395–396; 1999). "This bioinformatics initiative is a sort of test case for us, to see if we can be as flexible as we want the universities to be," says Engelke.

We normally disapprove of Cloning, Professor, but in this case...

BIOINFORMATICS DEPT.



Peer Bork, head of bioinformatics at the European Molecular Biology Laboratory in Heidelberg and a group leader at the Max Delbrück Centre for Molecular Medicine in Berlin, says he is "very pleased" by the style of the programme. Bork was a member of the group of experts called together last December by DFG president Ernst-Ludwig Winnacker to discuss the future of genomics research in Germany.

The group's deliberations formed the

basis of the DFG's position paper on genomics, which called on the government to provide an extra DM1 billion for genome research in the next five years (see *Nature* **399**, 622; 1999). The research ministry is now rethinking its genomics strategy and will publish a strategy paper in late summer. A major new programme in bioinformatics is expected to be announced.

But Bork warns that universities must take care when recruiting for any professorships created by the DFG programme. With few trained bioinformaticists available, and heavy competition from industry and other countries, "the new professorships could be filled with people with too little relevant experience: and this will be a problem for the training of the next generation".

These concerns are shared by Gerhardt Neuweiler, a former DFG president who is professor of zoology at the Ludwig Maximilian University in Munich, which is considering starting a course in bioinformatics. Neuweiler remembers when Germany "discovered ecology" 20 years ago. In the wave of the academic recruitment that followed, "many underqualified scientists were made professors, with the consequence that some areas of ecology are still very weak in Germany because of bad training".

There is general agreement that Germany, which has suffered a delayed start in all

areas of genomic research, needs to generate armies of bioinformatics graduates as quickly as possible. Jens Reich, for example, one of Germany's few dedicated bioinformaticists and a professor at Berlin's Humboldt University, says the dearth of young bioinformaticists "is very apparent".

All academic establishments say recruitment has been extremely difficult. Bork, for example, still has seven unfilled positions in his bioinformatics group — now 15-strong — because he has been unable to find suitably qualified candidates.

A lack of qualified scientists has also delayed the start of the DFG's special *Schwerpunkt* programme in bioinformatics, the participants for which were selected at the end of last year. University researchers on the programme could not recruit postdocs to do the work, and a few weeks ago, in desperation, the DFG advertised the programme abroad, and is now starting to recruit.

In contrast, industry, which offers higher salaries, has escaped this problem. Reinhardt Schneider, a founder of Lion Biosciences, Germany's biggest bioinformatics company, says he has had no trouble recruiting scientists. But only a minority of his bioinformaticists are German. He therefore welcomes the DFG initiative, although he adds that "it would have been even nicer if it had happened a few years earlier".

Alison Abbott

European embryology experts offer to advise on ethics of cloning

[MUNICH] The European Society for Human Reproduction and Embryology (ESHRE) last week confirmed its view that human cloning should not be used at present for reproductive purposes, but that the development of cloning techniques for therapeutics should not be banned.

The society, meeting in Tours, France, offered to advise politicians on cloning. Its president, Lynn Fraser, professor of reproductive medicine at King's College London, says the membership is broad enough to be able to give independent expert advice.

In the immediate future, she says, the society could help the British parliament decide whether to allow cloning techniques for therapeutic purposes. Britain's Human Fertilization and Embryology Authority recommended to parliament in December that these techniques should be allowed.

Last month, however, a parliamentary committee said it needed more time to consider and would seek additional advice. After the birth in 1997 of Dolly the sheep, the ESHRE called for a five-year moratorium on the reproductive cloning of humans. **A. A.**

Cable-car crash kills 20 radioastronomy workers in French Alps

[PARIS] A cable-car used to gain access to the Institute of Millimetric Radioastronomy (IRAM) facility in the French Alps crashed to the ground last week, killing all 20 people on board. The group was travelling to work at IRAM's radiotelescope observatory, 2,650 metres up a mountain slope, when the accident occurred. About 500 metres into the ride, the cable broke and the car plunged 80 metres to the ground.

IRAM is operated by France's Centre National de la Recherche Scientifique (CNRS), Germany's Max Planck Society and Spain's National Geographic Institute. The site, on top of the Bure plateau, hosts an array of five 15-metre telescopes, which allow the observation of interstellar molecules and dust from comets and stars.

An official investigation is being carried out into the cause of the accident, which is still unknown. The cable-car was the only link to the facility, so research has been severely hampered since the accident, with researchers relying on helicopters for access. IRAM was in the process of building a sixth radiotelescope, but its construction will now be delayed.

St Petersburg matches money from Soros

[MOSCOW] Officials in St Petersburg have started distributing individual research grants to match the money allotted to scientists by the International Educational Programme in Exact Sciences (ISSEP), which was set up by the international financier and philanthopist George Soros.

In doing so, Russia's second-largest city has joined 18 other Russian regions — including Moscow, Rostov, Samara and Tatarstan — in participating in the programme. The move was made after Vladimir Yakovlev, the governor of St Petersburg, met the city's scientists and educationalists. Over the past five years, Soros has put US\$76 million into ISSEP, while the Russian federal budget has given \$4 million and local authorities a further \$3 million.

France speeds up Internet access

[PARIS] French researchers logging onto the Internet this summer will notice a major increase in the speed of connections. The government announced last week an upgrade of the national research network, Renater, that will double its backbone speed

to 2.5 gigabits per second, with universities and private and public research units connected at up to 155 megabits per second.

The upgrade is seen as essential to stimulating the development of modelling, digital libraries, telemedicine and virtual campuses. Annual upgrades of about a further gigabit per second are also in the pipeline.

NIH makes large income from federal projects

[WASHINGTON] The US National Institutes of Health (NIH) dwarfs other federal agencies in terms of revenue received from the licensing of inventions based on federally funded research projects, according to a report by the General Accounting Office (see www.gao.gov/new.items/rc99173.pdf).

The report found that licences produced \$102.2 million in royalty revenue for the NIH, compared with just \$1.8 million, \$1.5 million and \$1 million for the Navy, the US space agency NASA and the Army, respectively, from 1996 to 1998. The Air Force and the Department of Energy together collected \$1 million.

Biosphere 2 to house carbon dioxide study

[WASHINGTON] The Packard Foundation has awarded a \$1 million grant to a research team that plans to use Biosphere 2, the giant greenhouse in southern Arizona, to study the effects on ecosystems of changing carbon dioxide levels.

Wallace Broecker, a geochemist at Columbia University in New York who was instrumental in the university taking over the management of the often-troubled facility three years ago, will lead a team of researchers in conducting the five-year project. The grant is the largest yet attracted by Biosphere 2.

A case of tuberculosis goes missing in hotel

[SAN DIEGO] A vial of infectious tuberculosis bacteria went missing last week in San Francisco when a thief stole luggage containing the sample from the hotel room of a researcher from the University of California at San Diego.

The researcher had just received the canister of bacteria from a colleague at the University of California at Berkeley during a conference on tuberculosis and leprosy. University officials say that all the appropriate procedures were followed in the handling of the vial, and the incident shows that researchers should guard samples closely while they are being transferred for studies.

Gore pledges to double cancer research funds

[WASHINGTON] US presidential candidate Al Gore said last week that, if elected, he would commit to doubling the amount that the federal government spends on cancer research over five years.

In a speech in Philadelphia in which he urged "a major new commitment to federal cancer research", the vice-president claimed that such an investment would save 200,000 lives and prevent 700,000 cases of cancer in the year 2010 alone. Gore also said that the commitment to cancer should be "matched by a similar commitment to all biomedical research".

ESA and John Innes Centre name new heads

[LONDON] The council of the European Space Agency (ESA) has a new chairman. Alain Bensoussan, president of the French Space Agency, was elected to a two-year term at the end of last month. He takes over from Hugo Parr, director general in the Ministry of Trade and Industry of Norway, whose three-year term of office ended on 30 June. Bensoussan, an engineer with a doctorate in mathematics, has been a member of the ESA council since 1996.

Meanwhile Christopher Lamb, currently regius professor of plant science at the University of Edinburgh, has been named as director of the John Innes Centre in Norwich, one of Europe's leading centres for plant science. Lamb spent much of his career at the Salk Institute, La Jolla, California, where he founded and directed the plant biology department. He takes over from Mike Gale, who has held the job for the past nine months. Gale will stay on as a senior member of staff at the institute.

Early American bones to be tested for age

[WASHINGTON] The controversial 9,000-year-old skeletal remains of a Native American found in Washington state three years ago will require more study, including destructive tests, to determine precisely the age of the bones.

The US Department of the Interior, which is embroiled in a legal battle over whether scientists should be given access to what some Native American groups consider to be sacred remains (see *Nature* 397, 551; 1999), will consult with Native American tribes before conducting further tests. Non-destructive analysis over the past few months has shown that the bones are all from one man, who was between 45 and 55 years old when he died, and that the skeleton is likely to have been buried intentionally.

Tests for BSE evaluated

Sir— Four rapid tests for the diagnosis of bovine spongiform encephalopathy (BSE) in bovines have been evaluated by us for the European Commission¹. Full details of this evaluation can be found in ref. 2.

We summarize here our findings on the sensitivity, specificity and detection limits of each test. Of ten tests submitted, four were accepted for evaluation on grounds such as their ability to be scaled up quickly for rapid evaluation on the timescale required. The tests were:

- **Test A (E. G. & G. Wallac)**: a two-site non-competitive immunometric procedure using two different monoclonal antibodies. DELFIA technology is used to generate the reading signal.
- **Test B (Prionics)**: an immunoblotting test based on a western blotting procedure for the detection of the protease-resistant fragment PrP^{Sc} using a monoclonal antibody.
- **Test C (Enfer)**: a chemiluminescent ELISA, using a polyclonal anti-PrP antibody for detection.
- **Test D (CEA)**: a sandwich immunoassay for PrP^{Sc} carried out following denaturation and concentration steps. Two monoclonal antibodies are used.

Tests A and D take less than 24 hours to perform, whereas tests B and C take 8 and 4 hours, respectively, and also have the highest throughput.

We evaluated sensitivity and specificity in relation to samples from true positive and true negative animals. We obtained positive samples from cows showing clinical signs of BSE and in which the disease was confirmed by histopathological examination. Negative

Table 1 Sensitivity and specificity using predetermined cut-off points

	Test A	Test B	Test C	Test D
Sensitivity	70%	100%	100%	100%
Specificity	90%	100%	100%	100%

samples were obtained from healthy cows of similar age slaughtered in New Zealand. The tissues used in the evaluation were from the same animals but were those tissues for which each test was developed: brain stem for three tests and anterior cervical spinal cord for the fourth, with a total of 1,000 negative and 300 positive samples for each test. Samples that weighed approximately 1g each were prepared, taking precautions to avoid any cross-contamination. In addition, positive brain homogenate of known infectivity titre was tested at dilutions in negative brain of up to 10⁻⁵ to estimate the detection limits of the tests.

We carried out the evaluation under supervision at the participants' laboratories over a one-month period. Testing was done with all involved blinded, including the supervisor being unaware of the identity of the samples. We interpreted the results using a cut-off point proposed by the participants. Inconclusive categories were established in advance and it was decided that, in the case of a retest of these samples, the second result would be the valid result. Results are summarized in Tables 1 and 2.

The results indicate that tests B, C and D have excellent potential for detecting or confirming clinical BSE for diagnostic purposes or for screening dead or slaughtered animals for such cases, particularly casualty animals or carcasses

Table 2 Number of homogenate samples scoring positive (above cut-off) at each dilution level

Dilution	Test A	Test B	Test C	Test D
0	6/6	6/6	6/6	6/6
10 ⁻¹	0/20	15/20 (+2?*)	20/20	20/20
10 ^{-1.5}		0/20	20/20	20/20
10 ^{-2.0}			0/20	20/20
10 ^{-2.5}				18/20
10 ^{-3.0}				1/20
10 ^{-3.5}				0/20

* Two samples rated inconclusive at this dilution.

sent for rendering. Even though BSE is a rare disease, the high specificity indicates that these tests may be useful for general post-mortem screening of older bovines.

The ability of tests to detect small concentrations of PrP^{Sc} gives grounds for optimism that they could detect infected animals before the development of clinical signs. However, the absence of information on the progression of the disease in bovines, particularly the relationship between infectivity titres and PrP^{Sc} concentration throughout the incubation period, means that it is not possible to reach any definite conclusions at this stage.

Jim Moynagh*, Heinz Schimmel†

*Directorate General XXIV, Consumer Policy and Health Protection, European Commission, rue Belliard 232, 1049 Bruxelles, Belgium

†Joint Research Centre, Institute for Reference Materials and Measurements, Retieseweg, B2440 Geel, Belgium

1. Butler, D. *Nature* **395**, 205–206 (1998).

2. European Commission *The Evaluation of Tests for the Diagnosis of Transmissible Spongiform Encephalopathies in Bovines* <http://europa.eu.int/comm/dg24/health/> & <http://www.irmm.jrc.be>

The land of rising science

Sir— Japan now has a greater scientific output than the United Kingdom, judging by the 1997 Japanese white paper (consultation document) on science and technology¹. The figures for 1995 clearly show that Japan ranks second only to the United States² in terms of scientific output. In my opinion, the Japanese scientific contribution is underestimated in the West.

Compared with the United Kingdom, Japan has a larger population (by a factor of 2.1), a larger economy (by a factor of 4.4), spends more on scientific research (by a factor of 3.5), has a larger base of researchers (by a factor of 3.7), and awards more advanced science degrees (by a factor of 1.9) (see Table 1 in Supplementary information and ref. 3).

Japanese industry spends relatively more on research and development, whereas Japanese government institutions and universities receive less funding, than their UK counterparts. Although Japan is officially in recession, these differences are so large that Japan continues to be ahead of the United Kingdom in these categories. Science funding in Japan has continued to increase despite economic difficulties.

Japan has a 23.9% share of world exports of high-tech products, compared with the 8.3% UK share. It has a trade balance (exports/imports) of 3.52 for high-tech products compared with 0.93 for the United Kingdom.

These figures show unequivocally that Japan has a much bigger scientific base than the United Kingdom and that Japanese companies are much more engaged in research than their UK counterparts. These factors would, in part, explain Japan's dominance in technology-based industries.

In 1994, Japan produced 9.6% of the world's scientific papers published in major journals and had an 8.0% share of the number of citations of papers (see Tables 2 and 3 in Supplementary information). Corresponding figures for the United Kingdom were 9.1% and 11.6%, respectively. This undoubtedly results from the much greater increase in research and development expenditure in Japan (237%) than in the United Kingdom (126%) during 1980–95.

A fuller version of this Correspondence is at <http://www.gsj.go.jp/~kiyo/glasby.html>

G. P. Glasby

42 Warminster Crescent, Sheffield S8 9NW, UK (correspondence address); and Earthquake Research Institute, University of Tokyo, Tokyo, Japan

1. *White Paper on Science and Technology 1997* (Japan Sci. and Tech. Corp., Tokyo, 1997). (http://www.sta.go.jp/shokai/publications/1997_white_paper/JMAH.HTM)

2. May, R. M. *Science* **275**, 793–796 (1997).

3. Wisniewski, D. *Annual Abstracts of Statistics 1998* Edn No. 134 (HMSO, London, 1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Accessibility of information on the web

Search engines do not index sites equally, may not index new pages for months, and no engine indexes more than about 16% of the web. As the web becomes a major communications medium, the data on it must be made more accessible.

Steve Lawrence and C. Lee Giles

The publicly indexable World-Wide Web now contains about 800 million pages, encompassing about 6 terabytes of text data on about 3 million servers. The web is increasingly being used in all aspects of society; for example, consumers use search engines to locate and buy goods, or to research many decisions (such as choosing a holiday destination, medical treatment or election vote). Scientists are increasingly using search engines to locate research of interest: some rarely use libraries, locating research articles primarily online; scientific editors use search engines to locate potential reviewers. Web users spend a lot of their time using search engines to locate material on the vast and unorganized web. About 85% of users use search engines to locate information¹, and several search engines consistently rank among the top ten sites accessed on the web².

The Internet and the web are transforming society, and the search engines are an important part of this process. Delayed indexing of scientific research might lead to the duplication of work, and the presence and ranking of online stores in search-engine listings can substantially affect economic viability (some websites are reportedly for sale primarily based on the fact that they are indexed by Yahoo).

We previously estimated³ that the publicly indexable web contained at least 320 million pages in December 1997 (the publicly indexable web excludes pages that are not normally considered for indexing by web search engines, such as pages with authorization requirements, pages excluded from indexing using the robots exclusion standard, and pages hidden behind search forms). We also reported that six major public search engines (AltaVista, Excite, HotBot, Infoseek, Lycos and Northern Light) collectively covered about 60% of the web. The largest coverage of a single engine was about one-third of the estimated total size of the web.

How much information?

We have now obtained and analysed a random sample of servers to investigate the amount and distribution of information on the web. During 2–28 February 1999, we chose random Internet Protocol (IP) addresses, and tested for a web server at the standard port. There are currently 256⁴ (about 4.3 billion) possible IP addresses (IPv6, the next version of the IP protocol

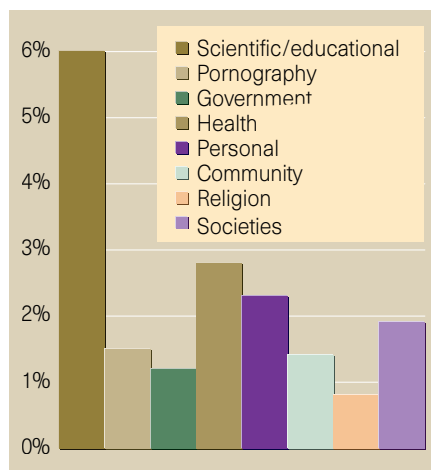


Figure 1 Distribution of information on the publicly indexable web as of February 1999. About 83% of servers contain commercial content (for example, company home pages). The remaining classifications are shown above. Sites may have multiple classifications.

which is under development, will increase this substantially); some of these are unavailable while some are known to be unassigned. We have tested random IP addresses (with replacement), and have estimated the total number of web servers using the fraction of tests that successfully locate a server. Many sites are temporarily unavailable because of Internet connectivity problems or web-server downtime, and to minimize this effect, we rechecked all IP addresses after a week.

Testing 3.6 million IP addresses (requests timed out after 30 seconds of inactivity), we found a web server for one in every 269 requests, leading to an estimate of 16.0 million web servers in total. For comparison, Netcraft found 4.3 million web servers in February 1999 based on testing known host names (aliases for the same site were considered as distinct hosts in the Netcraft survey (www.netcraft.com/survey/)). The estimate of 16.0 million servers is not very useful, because there are many web servers that would not normally be considered part of the publicly indexable web. These include servers with authorization requirements (including firewalls), servers that respond with a default page, those with no content (sites 'coming soon', for example), web-hosting companies that present their home page on many IP addresses, printers, routers, proxies, mail servers, CD-ROM servers and other hardware that provides a web interface. We built a database of regular expressions to identify most of these servers. For the results

reported here, we manually classified all servers and removed servers that are not part of the publicly indexable web. Sites that serve the same content on multiple IP addresses were accounted for by considering only one address as part of the publicly indexable web.

Our resulting estimate of the number of servers on the publicly indexable web as of February 1999 is 2.8 million. Note that it is possible for a server to host more than one site. All further analysis presented here uses only those servers considered part of the publicly indexable web.

To estimate the number of indexable web pages, we crawled all the pages on the first 2,500 random web servers. The mean number of pages per server was 289, leading to an estimate of the number of pages on the publicly indexable web of about 800 million. It is important to note that the distribution of pages on web servers is extremely skewed, following a universal power law⁴. Many sites have few pages, and a few sites have vast numbers of pages, which limits the accuracy of the estimate. The true value could be higher because of very rare sites that have millions of pages (for example, GeoCities reportedly has 34 million pages), or because some sites could not be crawled completely because of errors.

The mean size of a page was 18.7 kilobytes (kbytes; median 3.9 kbytes), or 7.3 kbytes (median 0.98 kbytes) after reducing the pages to only the textual content (removing HTML tags, comments and extra white space). This allows an estimate of the amount of data on the publicly indexable web: 15 terabytes (Tbytes) of pages, or 6 Tbytes of textual content. We also estimated 62.8 images per web server and a mean image size of 15.2 kbytes (median 5.5 kbytes), leading to an estimate of 180 million images on the publicly indexable web and a total amount of image data of about 3 Tbytes.

We manually classified the first 2,500 randomly found web servers into the classes shown in Fig. 1. About 6% of web servers have scientific/educational content (defined here as university, college and research lab servers). The web contains a diverse range of scientific material, including scientist, university and project home pages, preprints, technical reports, conference and journal papers, teaching resources, and databases (for example, gene sequences, molecular structures and image libraries). Much of this material is not available in traditional databases. Research articles that do become available in traditional databases are often available earlier on the web⁵ (for example,

scientists may post preprints on their home pages). The high value of the scientific information on the web, and the relatively small percentage of servers that contain the bulk of that information, suggest that an index of all scientific information on the web would be feasible and very valuable.

We also analysed metadata on the home pages of each server using the HTML META tag. Many different META tags are used, most of which encode details that do not identify the content of the page. We used the existence of one or more of the keywords or description tags as evidence that effort was put into describing the content of the site. We found that 34.2% of servers contain such metadata on their home page. The low usage of the simple HTML metadata standard suggests that acceptance and widespread use of more complex standards, such as XML or Dublin Core⁶, may be very slow (0.3% of sites contained metadata using the Dublin Core standard). We also noted great diversity in the HTML META tags found, with 123 distinct tags, suggesting a lack of standardization in usage.

Search engine performance

We previously provided coverage results for six major web search engines in December 1997 (ref. 3). We provide updated results here for February 1999 using 11 major full-text search engines (in alphabetical order): AltaVista, EuroSeek, Excite, Google, HotBot, Infoseek, Lycos, Microsoft, Northern Light, Snap and Yahoo. We analysed the Yahoo database provided by Inktomi, not its hierarchical taxonomy (which is much smaller). HotBot, Microsoft, Snap and Yahoo all use Inktomi as a search provider, but the results differ between these engines (due to filtering and/or different underlying Inktomi databases). We did not include Northern Light's 'special collection' (documents that are not part of the publicly indexable web).

We performed our study by analysing the responses of the search engines to real-world queries, as we are interested in the relative coverage of the engines for real queries, which can be substantially different from the relative number of pages indexed by each engine. The engines differ in the specific

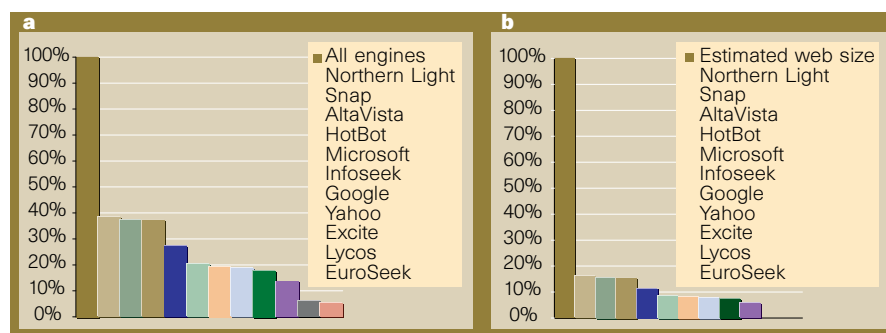


Figure 2 Relative coverage of the engines for the 1,050 queries used during the period 25–28 February 1999. Note that these estimates are of relative coverage for real queries, which may differ from the number of pages actually indexed because of different indexing and retrieval techniques. a, Coverage with respect to the combined coverage of all engines; b, coverage with respect to the estimated size of the web.

details of indexing and retrieval, for example there may be different maximum limits on the size of pages indexed by each engine, restrictions on which words are indexed or limits on processing time used per query. This means that comparison has to be done on real queries; comparing the size of the engine databases (readily available from the engines) would lead to misleading results.

In our previous study³, we used 575 queries originally performed by the employees of NEC Research Institute (mostly scientists). We retrieved the entire list of documents for each query and engine, and then retrieved every individual document for analysis. Only documents containing all of the query terms are included because the search engines can return documents that do not contain the query terms (for example, documents with morphological variants or related terms). These documents may be relevant to the query, but they prevent accurate comparison across search engines. In this study, we have increased the number of search engines analysed to 11, expanded the number of queries to 1,050, and transformed queries to the advanced syntax for AltaVista which allows retrieval of more than 200 results.

We used automatic and manual consistency checks to ensure that responses from the search engines were being correctly processed. The relative coverage of each engine is shown in Table 1 and Fig. 2. Note

that Lycos and HotBot returned only one page per site (for Infoseek we used the option not to group results by site). Northern Light has increased substantially in coverage since December 1997, and is now the engine with the greatest coverage (for our queries).

To express the coverage of the engines with respect to the size of the web, we need an absolute value for the number of pages indexed by one of the engines. We used 128 million pages for Northern Light, as reported by the engine itself at the time of our experiments. Our results show that the search engines are increasingly falling behind in their efforts to index the web (Fig. 2).

The overlap between the engines remains relatively low; combining the results of multiple engines greatly improves coverage of the web during searches. The estimated combined coverage of the engines used in the study is 335 million pages, or 42% of the estimated total number of pages. Hence, a substantial improvement in web coverage can be obtained using metasearch engines, such as MetaCrawler⁷, which combine the results of multiple engines.

When searching for 'not x', where 'x' is a term that does not appear in the index, AltaVista (Boolean advanced search with the 'count matching pages' option) and Northern Light report the total number of pages in their index. Figure 3 shows the reported number of pages indexed by these engines between 1 November 1998 and 1 April 1999.

Table 1 Statistics for search-engine coverage and recency

Search engine	Northern Light	Snap	AltaVista	HotBot	Microsoft	Infoseek	Google	Yahoo	Excite	Lycos	EuroSeek	Average
Coverage with respect to combined coverage (%)	38.3	37.1	37.1	27.1	20.3	19.2	18.6	17.6	13.5	5.9	5.2	n/a
95% confidence interval	±0.82	±0.75	±0.77	±0.64	±0.51	±0.82	±0.69	±0.61	±0.46	±0.30	±0.40	n/a
Coverage with respect to estimated web size	16.0	15.5	15.5	11.3	8.5	8.0	7.8	7.4	5.6	2.5	2.2	n/a
Percentage of invalid links	9.8	2.8	6.7	2.2	2.6	5.5	7.0	2.9	2.7	14.0	2.6	5.3
Mean age of new matching documents (days)	141	240	166	192	194	148	n/a	235	206	174	n/a	186
Median age of new matching documents (days)	84	91	33	51	57	60	n/a	76	47	174	n/a	57

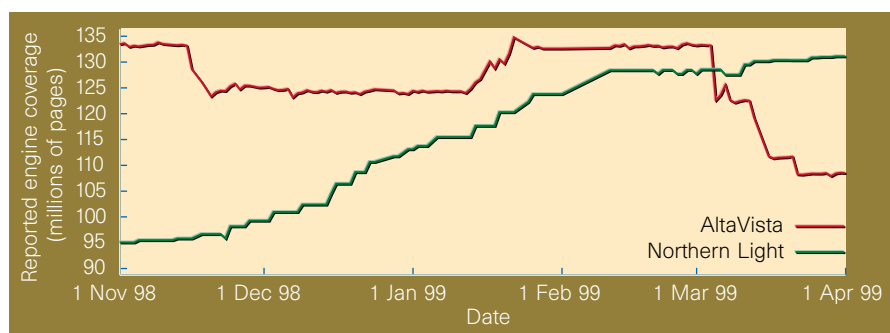


Figure 3 Number of pages indexed by AltaVista and Northern Light between 1 November 1998 and 1 April 1999 (sampled daily), as reported by the engines themselves.

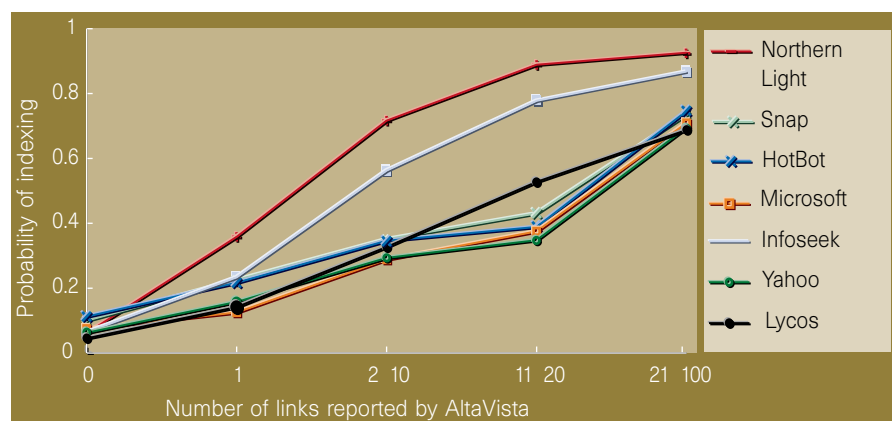


Figure 4 Probability of indexing random sites versus 'popularity' or connectedness of sites. Sites with few links to them have a low probability of being indexed.

The number of reported pages indexed by AltaVista has decreased significantly at times, suggesting possible difficulty in maintaining an index of more than 100 million pages.

We next looked at statistics that indicate how up to date the search-engine databases are. Table 1 shows the results of two investigations for each search engine. The first investigation looks at the percentage of documents reported by each engine that are no longer valid (because the page has moved or no longer exists). Pages that timed out are not included. The indexing patterns of the search engines vary over time, and the engine with the most up-to-date database may not be the most comprehensive engine.

We have also looked at the age of new documents found that match given queries by tracking the results of queries using an internal search engine. Queries were repeated at the search engines daily, and new pages matching the queries were reported to the searchers. We monitored the queries being tracked, and when new documents were found, we logged the time since the documents were last modified (when available), which provides a lower bound for the time span between when documents matching a query are added or modified, and the time that they are indexed by one of the search engines (new queries are not included in the study for the first week). The mean age of a matching page when indexed by the first of the nine engines is 186 days, while the medi-

an age is 57 days (these averages are over all documents, not over search engines) (Table 1). Although our results apply only to pages matching the queries performed and not to general web pages, they do provide evidence that indexing of new or modified pages can take several months or longer.

Why do search engines index such a small fraction of the web? There may be a point beyond which it is not economical for them to improve their coverage or timeliness. The engines may be limited by the scalability of their indexing and retrieval technology, or by network bandwidth. Larger indexes mean greater hardware and maintenance expenses, and more processing time for at least some queries on retrieval. Therefore, larger indexes cost more to create, and slow the response time on average. There are diminishing returns to indexing all of the web, because most queries made to the search engines can be satisfied with a relatively small database. Search engines might generate more revenue by putting resources into other areas (for example, free e-mail). Also, it is not necessary for a page to be indexed for it to be accessible: a suitable query might locate another page that links to the desired page.

Not equal access

One of the great promises of the web is that of equalizing the accessibility of information. Do the search engines provide equal

access? They typically have two main methods for locating pages: following links to find new pages; and user registration. Because the engines follow links to find new pages, they are more likely to find and index pages that have more links to them. The engines may also use various techniques for choosing which pages to index. Therefore, the engines typically index a biased sample of the web.

To analyse bias in indexing, we measured the probability of a site being indexed as a function of the number of links to the site. We used AltaVista to obtain an estimate of the number of links to the home pages of the 2,500 random web servers found earlier. We then tested the existence of the sites in all the other engines that support URL search. Figure 4 shows the probability of each engine indexing a random site versus the number of links to the site. A very strong trend can be seen across all engines, where sites with few links to them have a low probability of being indexed and sites with many links to them have a high probability of being indexed. Considering the relative coverage results, Infoseek has a higher probability of indexing random sites than might be expected. This indicates that Infoseek may index the web more broadly (that is, index more sites instead of more pages per site).

Not only are the engines indexing a biased sample of the web, but new search techniques are further biasing the accessibility of information on the web. For example, the search engines DirectHit and Google make extensive use of measures of 'popularity' to rank the relevance of pages (Google uses the number and source of incoming links while DirectHit uses the number of times links have been selected in previous searches). The increasing use of features other than those directly describing the content of pages further biases the accessibility of information. For ranking based on popularity, we can see a trend where popular pages become more popular, while new, unlinked pages have an increasingly difficult time becoming visible in search-engine listings. This may delay or even prevent the widespread visibility of new high-quality information. □

Steve Lawrence and C. Lee Giles are at NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA.
e-mail: lawrence@research.nj.nec.com

1. Graphic, Visualization, and Usability Center. *GUV's 10th WWW User Survey* (October 1998); http://www.gvu.gatech.edu/user_surveys/survey-1998-10/
2. Media Metrix (January 1999); http://www.mediametrix.com/PressRoom/Press_Releases/02_22_99.html
3. Lawrence, S. & Giles, C. L. *Science* **280**, 98–100 (1998).
4. Huberman, B. A. & Adamic, L. A. *Evolutionary Dynamics of the World Wide Web* (1999); <http://www.parc.xerox.com/istl/groups/iea/www/growth.html>
5. Lawrence, S., Giles, C. L. & Bollacker, K. *IEEE Computer* **32**(6) (1999).
6. Dublin Core. *The Dublin Core: A Simple Content Description Model for Electronic Resources* (1999); <http://purl.oclc.org/dc/>
7. Selberg, E. & Etzioni, O. *IEEE Expert* 11–14, Jan.–Feb. (1997).

Less expansion, more agreement

Riccardo Giovanelli

We have known for 70 years that the Universe is expanding, but astronomers couldn't agree on how fast. We may be approaching – with 10% uncertainty – a consensus value. So is one of astronomy's greatest quests finally over?

Edwin Hubble got it wrong by an order of magnitude¹, but claimed an uncertainty of 10%. Seventy years and much humble pie later, our knowledge of the value of the current rate of expansion of the Universe, H_0 , may be approaching that degree of precision. At the American Astronomical Society meeting last month*, members of the Hubble Space Telescope Key Project team announced the results of their eight-year search for H_0 .

The determination of H_0 requires independent measurements of the recession velocities and distances of extragalactic sources. The relative speeds of distant objects are routinely obtained with high precision from Doppler shifts of spectral lines of stellar or interstellar origin; distance measurements are difficult and more uncertain. The first estimates of H_0 given by Edwin Hubble (Fig. 1) hovered near $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$. A megaparsec (Mpc) is a unit of distance equivalent to 3.1 million light years; two galaxies separated by $d \text{ Mpc}$ recede from each other at $H_0 d \text{ km s}^{-1}$. These ungainly but practical units used by astronomers will be assumed, but explicitly omitted, throughout this article. Had the expansion of the Universe proceeded unchanged at the galloping rate suggested by Hubble, the Big Bang would have taken place less than two billion years ago, making the Earth older than the Universe. In the 1950s, progress in stellar astronomy and the recognition of observational blunders brought estimates of H_0 down to less than 100, and there it long stayed, tugged between 50 and 100 by raspy conflict for three decades.

The reasons for such large uncertainty stem from the uneasy set of extrapolations required by distance measurements. The operational approach to measuring distances is referred to as the 'cosmological distance ladder', which splices together techniques of progressively greater scope until distances of cosmological interest are reached. With the advent of the Hubble Space Telescope (HST), the opportunity arose to substantially improve the distances to which one of the most important links of the ladder could be applied. Cepheids are

bright, regularly variable stars (the period of their variability is correlated with their mean luminosity), which are assumed for the purposes of the distance ladder to be 'standard candles'. Observations of their apparent brightness at different times indicate their distances, and that of their parent galaxies, from Earth. Several hundred Cepheids have been found with the HST, at distances of a few tens of Mpc, mainly by a team led by W. Freedman, R. Kennicutt and J. Mould (Key Project group²). These have provided a much surer foothold on the distance ladder.

For a reliable value of H_0 , the distance ladder needs to be extended farther than the largest distances at which Cepheids can be seen, so that perturbations produced by mass-density fluctuations in the local Universe have a negligible influence on estimates of Hubble expansion. The next rungs of the ladder are provided by a variety of techniques that more often rely on the relationship between rotational speed and luminosity of spiral galaxies, or the surface brightness fluctuations in the images of elliptical galaxies, or the decay in the light emission of supernovae explosions. In the past, these methods generally yielded significantly different values of H_0 , with the lowest estimate coming from supernovae. But the gap is narrowing. Results from three studies by the HST Key Project team give values of $H_0 = 71$ from spiral galaxies (S. Sakai), $H_0 = 69$ from elliptical galaxies (L. Ferrarese) and $H_0 = 68$ from supernovae (B. Gibson). They all reported uncertainties near 10%.

Other groups have used the Cepheid calibration data in different ways. Another supernovae study found $H_0 = 60$ (A. Saha, Carnegie Observatories), whereas $H_0 = 73\text{--}79$ was obtained from elliptical data (J. Tonry, Univ. Hawaii), and $H_0 = 77$ was calculated from spirals (B. Tully, Univ. Hawaii and M. Pierce, Indiana Univ.). Another spiral study by a French–Finnish group³ has reported $H_0 \approx 55$. The source of the discrepancies is to be found among uncertainties in the corrections for sampling biases³ and in the deviations from smooth Hubble expansion in the local Universe, which can affect inferences of H_0 from nearby sources by several per cent^{4,5}. Furthermore, differences in the chemical compositions of Cepheids,

which can produce shifts in the period–luminosity relation⁶, may lead to systematic errors in distance estimates. So, error assessments are still loaded with unknown systematic components. This is well illustrated by the report of a trigonometric distance to the galaxy NGC4258 by radio interferometry (J. Herrnstein, Nat. Radio Astronomy Observatory, New Mexico); it differs by about 15% from the Cepheid distance to the same galaxy, whereas reported errors of both techniques vary between 4% and 7%. As systematic effects are better understood, their impact diminishes and the gap between various estimates of H_0 narrows. At this time, our understanding of these errors seems to hold H_0 at the limit of, but not below, the threshold of 10% accuracy.

In the past few years, new sets of distance indicators, which do not rely on cumbersome distance-ladder extrapolations, have delivered increasingly accurate results. The Sunyaev–Zeldovich effect (SZE) depends on a spectral shift in the photons of the cosmic microwave background, produced by their interaction with the hot gas dwelling in the core of galaxy clusters. A joint measurement of the SZE shift and the X-ray emission of intracluster gas yields an estimate of a cluster's distance. Groups using this technique find values of H_0 ranging from 71 (B. Mason and S. Myers, Univ. Pennsylvania, Philadelphia), to between 57 and 64 (J. Carlstrom, Univ. Chicago). Uncertainties in H_0 from SZE determinations are about 15%, arising



Figure 1 Edwin Hubble, who discovered that the Universe is expanding. Knowing the rate of expansion, the Hubble constant, is essential for determining the age and size of the Universe.

*American Astronomical Society Centennial Meeting, Chicago, Illinois, USA, 30 May to 3 June, 1999.



100 YEARS AGO

The Science Section of the Women's Congress was held at the Small Hall in the Westminster Town Hall on Thursday, June 29 ... The subject of research work was discussed, and stress was laid upon the fact that, inasmuch as the majority of students who take up science do so either as an avenue to a degree or with the idea of earning a livelihood by teaching later on, their training was as a rule insufficient and quite inadequate to permit them to undertake independent original work; whilst on the other hand the demands upon their time made by teaching was so great as to leave practically no leisure for higher work, even when they were qualified to do it. ... It was highly satisfactory to find that, in the open discussion which followed, an attempt on the part of two speakers to introduce the question of vivisection from the anti-vivisectionist point of view was not tolerated by the audience, these speakers being refused a hearing. It is not too much to say that the papers contributed were worthy both of their subjects and their authors, and that there was a refreshing absence of the hackneyed comparison of the relative position and intellectual powers of men and women, which has been such a favourite theme with so many speakers at this Congress.

From *Nature* 6 July 1899.

50 YEARS AGO

Dr. E. Rosen has written a remarkably lively little book, the conclusion of which is that the term 'telescope' was originally devised by John Demisiani of Cephalonia, and made public by Frederick Cesi at the banquet given in honour of Galileo on April 14, 1611. The circumstances may seem to the man of science of to-day to be too minute to warrant the publication of a book; but we must welcome here what the history of science so notably lacks, namely, a painstaking and detailed inquiry into a doubtful point concerning which both contemporaries and historians are at variance. The book scarcely needs to justify itself by its subject-matter, for such is the gusto of its author that the reader is infected by that strange passion that keeps sensible men out of bed after midnight to read detective stories.

From *Nature* 9 July 1949.

principally from the imperfectly known geometry of the clusters. Because SZE can be measured on very distant clusters, the technique samples smooth Hubble expansion (negligibly disturbed by local effects), but its results are dependent on the large-scale geometry of the Universe, which accounts for the spread in the values reported by Carlstrom. An alternative technique that measures time delays between multiple images of a distant source (produced by the gravitational lensing of an intervening mass concentration) is also starting to yield interesting results (P. Schechter, MIT). Findings from a study of lens Q0957 + 561 gives $H_0 = 61$, with an uncertainty of 20% (A. Romanowsky and C. Kochanek)⁷.

In matter-dominated cosmological models, values of H_0 above 60 have the embarrassing feature of yielding an age for the Universe since the Big Bang that is exceeded by the oldest stars in our Galaxy, which are approaching an elderly 12 billion years⁸. With the discovery^{9,10} that the expansion of the Universe may be accelerated by dark, perhaps vacuum energy, even relatively high values of H_0 become consistent with stellar astrophysics,

removing this age paradox. Since the 1950s, Allan Sandage in particular has towered over efforts to measure H_0 . At times it seemed like Sandage and Gustav Tammann were holding a tug-of-war at the 60 km s⁻¹ Mpc⁻¹ line, against a swelling crowd pulling for the 80s, 90s and even higher. After several decades of taking this heroic stance, it appears to me that Sandage and Tammann have a fair claim to victory — it may not be 43 (ref. 11), but 66 is starting to look pretty good. □

Riccardo Giovanelli is in the Department of Astronomy, Cornell University, Ithaca, New York 14853, USA.

e-mail: riccardo@astrosun.tn.cornell.edu

1. van den Bergh, S. in *The Extragalactic Distance Scale* (eds Livio, M., Donahue, M. & Panagia, N.) 1 (Cambridge Univ. Press, 1997).
2. <http://oposite.stsci.edu/pubinfo/pr/1999/19/key.html>
3. Ekholm, T. *et al.* <http://xxx.lanl.gov/astro-ph/9904372>
4. Zehavi, I., Riess, A. G., Kirshner, R. P. & Dekel, A. *Astrophys. J.* **503**, 483–491 (1998).
5. Giovanelli, R., Dale, D., Haynes, M. P., Hardy, E. & Campusano, L. *Astrophys. J.* (in the press).
6. Kennicutt, R. *et al.* *Astrophys. J.* **498**, 181–194 (1998).
7. Romanowsky, A. & Kochanek, C. *Astrophys. J.* **516**, 18–36 (1999).
8. Chaboyer, B., Demarque, P., Kernan, P. J. & Krauss, L. M. *Astrophys. J.* **494**, 96–110 (1998).
9. Perlmutter, S. *et al.* *Astrophys. J.* **517**, 565–590 (1999).
10. Riess, A. G. *et al.* *Astron. J.* **116**, 1009–1038 (1998).
11. Sandage, A. *Astrophys. J.* **402**, 3–14 (1993).

Aerodynamics

How flies fly

Robin Wootton

Insects are by far the oldest, most numerous and smallest flying machines. Writing in *Science*, Dickinson and colleagues¹ now fill a large gap in our understanding of their unconventional — but effective — aerodynamic techniques. Using a flapping, robotic fruitfly (*Drosophila*), Dickinson *et al.* show how transient bursts of lift can be generated by the wings as they rotate at the top and bottom of each stroke. Their work not only contributes to this lively field of study, but goes a long way towards providing a general theory of insect flight.

Like helicopters, but unlike conventional aircraft, insects support their weight, propel themselves and control their attitude and manoeuvres by actively moving their wings relative to their body and accelerating a precisely directed mass of air: directly down if hovering; obliquely down if flying forwards, backwards or sideways. Helicopters rotate their wings around a more-or-less vertical shaft. Insects, which have no rotary bearings, oscillate and twist their wings, generally ensuring that as much of the stroke-cycle as possible is creating usefully directed aerodynamic force.

Conventional aerodynamic theory was built on fixed wings in a steady airflow. It has been plain for a quarter of a century that the airflow around insects is anything but steady, and that they make extensive use of aerodynamic tricks to gain remarkably high levels

of lift. In the past few years a good deal of research has been directed towards uncovering the mechanisms involved.

A flying insect beats its wings down relative to its body; then twists them backwards around their longitudinal axes; beats them up; and twists them forwards. Downstroke and upstroke — which are actually usually oblique, sometimes even horizontal — are the translational phases of the stroke. The brief periods of twisting at the top and bottom of the stroke are the rotational phases.

Wings generate lift by creating vortices: whirlwind-like circulations of air around a linear core. Crucial among these is the bound vortex, which circulates around the wing itself, speeding up the mass airflow over the upper surface of the wing, slowing it down beneath, accelerating it down behind, and creating a pressure difference between the upper and lower surfaces which tends to suck the wing upwards. The stronger this circulation, the higher the lift; and both increase with the angle of attack — the angle between the wing and the oncoming flow — until the critical stalling angle is reached, when the flow abruptly separates from the upper wing surface and lift disappears.

If a wing is suddenly accelerated at a high angle of attack, stalling can briefly be delayed, and extraordinarily high lift generated, by the formation of another vortex on the upper surface of the wing just behind the

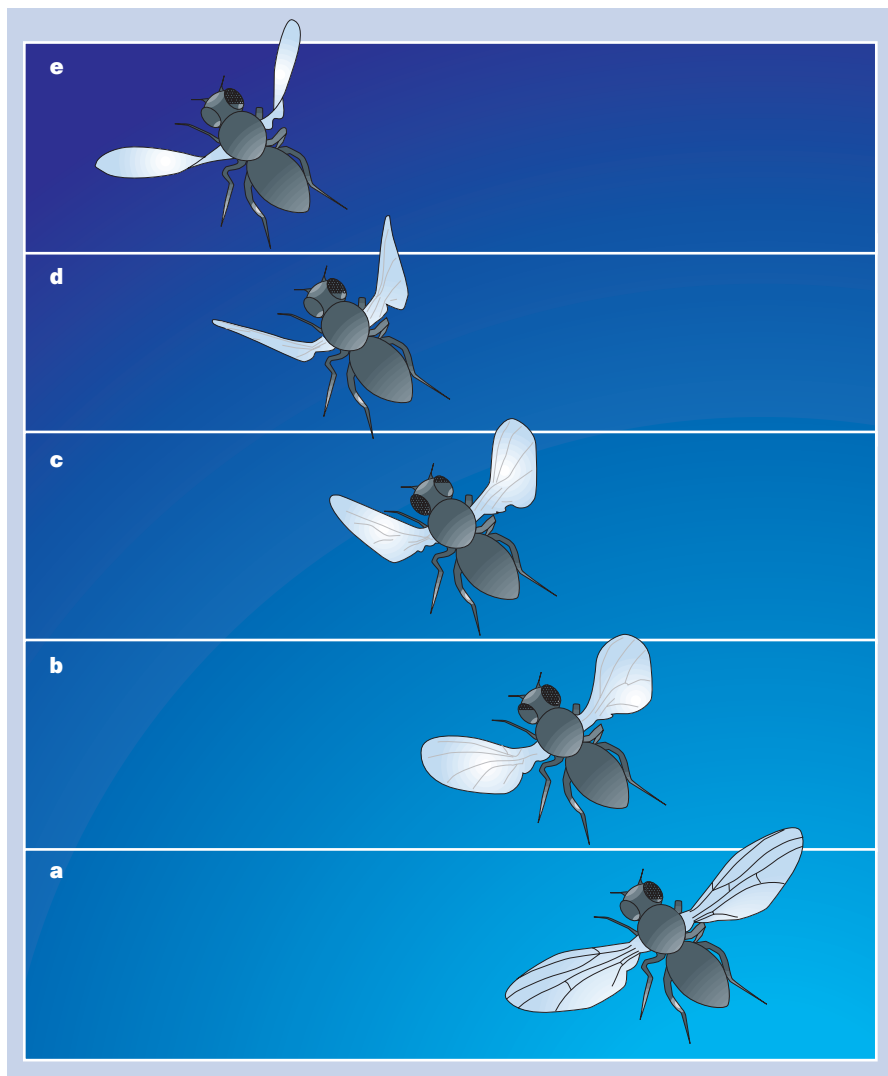


Figure 1 Wing twisting around the end of the downstroke in *Drosophila*. Here a is around the middle of the downstroke, d is the bottom, and in e the wing is entering the upstroke. Backwards rotation clearly begins (c) before the wing reaches the bottom of the stroke, and Dickinson *et al.*¹ show that it generates a pulse of high aerodynamic lift towards the end of this half of the cycle. As the strongly twisted wing begins to move upwards and backwards (e) a second pulse of high lift is created, apparently by the wings intercepting and capturing the remains of the downstroke wake. The sequence shown here is traced from high-speed cine film of a freely flying insect taken by A. R. Ennos.

leading edge, which causes the flow to re-attach to the surface instead of separating from it. This is a momentary phenomenon, and was assumed to be too short-lived to be significant in flapping flight — until 1996, that is, when Ellington *et al.*² demonstrated in the hawkmoth *Manduca* that the flapping motion of the wing stabilizes the leading-edge vortex by causing it to spiral out along the wing towards the tip. In this way, stalling is delayed and abnormally high lift maintained through the translatory phases of the stroke.

The effect was confirmed by a combination of stereophotography of the three-dimensional structure of the wakes of the hawkmoths themselves³; computational fluid-dynamic analysis, based on the observed wing movements⁴; and examination of the movement of smoke around the wings of a large robot hawkmoth precisely

copying the actual movements and deformations of the wings of the real moth, but with the flapping frequency scaled down to ensure the same flow pattern^{5,6}. Further investigations, including those of Dickinson's group¹, are confirming that delayed stall is the main source of extra lift in the translational phases of the upstroke and downstroke of a wide range of insects.

What, though, is happening as the wing twists forwards and backwards around the upper and lower extremes of the stroke? To tackle this question, Dickinson *et al.* used another robot, with Plexiglas *Drosophila*-shaped wings instrumented to measure lift and drag, operating in a tank of oil to replicate the pattern of flow that the actual wings create in air. When the wings were flapped and rotated to simulate the *Drosophila* stroke cycle, brief bursts of high lift occurred just

before and just after the top and bottom of each stroke. When the pattern was modified experimentally, it became clear that the timing of rotation was crucial. The first peak in lift depends on the wing beginning to twist before the extreme point of each half-stroke, creating a high angle of attack as it is decelerating (Fig. 1). The effect is like giving spin to a tennis ball. The backflip of the wing just before the end of the downstroke accelerates the airflow past the upper side — backspin — generating the first burst of lift. The flip before the end of the oblique upstroke, effectively another backflip, does the same.

So much for the lift peak before each stroke reversal. The explanation offered for the peak that immediately follows is perhaps still more remarkable. The high flapping velocity of *Drosophila*, and its relatively low flight speed, ensure that after each reversal the wing passes back through the still-rotating air left behind by the previous half-stroke. Using a flow-visualization technique known as particle-imaging velocimetry, Dickinson and colleagues have found evidence that, again provided that the wing has rotated before it changes direction, it recaptures the wake that it has just shed, and uses its lingering vorticity to generate another burst of lift.

It has yet to be shown that these phenomena are widespread in flying insects. Nonetheless, they seem most satisfyingly to complement the leading-edge vortex story of Ellington's group in providing an explanation of lift generation throughout the entire wing-stroke cycle. The calculations of Dickinson and colleagues indicate that wing translation still supplies the greater part of the lift, but that the brief periods of wing twisting provide a significant component. Furthermore, the sensitivity of this component to the timing of wing twisting makes it ideally suited for the initiation and control of the sharp accelerations and manoeuvres that so characterize the flight of higher flies. Minor changes in the timing of the twisting at the top and bottom of the stroke in one or both wings would have major, near-instantaneous effects on speed and direction, amply demonstrated by high-speed film of flies in free flight⁷. No wonder they are difficult to swat. □

Robin Wootton is in the School of Biological Sciences, University of Exeter, Hatherly Laboratories, Prince of Wales Road, Exeter EX4 4PS, UK.

e-mail: R.J.Wootton@exeter.ac.uk

- Dickinson, M. H., Lehmann, F.-O. & Sane, S. P. *Science* **284**, 1954–1960 (1999).
- Ellington, C. P., van den Berg, C., Willmott, A. P. & Thomas, A. L. R. *Nature* **384**, 626–630 (1996).
- Willmott, A. P., Ellington, C. P. & Thomas, A. L. R. *Phil. Trans. R. Soc. Lond. B* **352**, 303–316 (1997).
- Liu, H., Ellington, C. P., Kawachi, K., van den Berg, C. & Willmott, A. P. *J. Exp. Biol.* **201**, 461–477 (1998).
- van den Berg, C. & Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **352**, 317–328 (1997).
- van den Berg, C. & Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **352**, 329–340 (1997).
- Ennos, A. R. *J. Exp. Biol.* **142**, 49–85 (1989).

Computing

Solving problems in finite time

Philip W. Anderson

I remember with pleasure and some surprise the two weeks that my high-school maths teacher took out of teaching advanced algebra to introduce us to the kind of logical puzzle where, for instance, you may be given: three last names, say Smith, Jones and Doe; three houses, painted white, red and blue respectively; and three first names, Alice, Barbara and Carol. You are told a carefully rationed set of facts — such as Alice had a quarrel with Mrs Smith and Carol admired her friend's blue house — and asked to connect first and last names and house colours. It turns out (although it is not part of high-school maths even today) that these are all special cases of one of the classic problems of computer science, the 'K-SAT' problem. This is a good example of the type of problem where the computer time required to arrive at a solution may explode exponentially with the number of variables, or the number of propositions one is asked to analyse. On page 133 of this issue, Monasson *et al.*¹ present a new analysis of the K-SAT problem that throws light on when and why such computing tasks become prohibitive.

The K-SAT problem was the first to be assigned to the class of NP (nondeterministic polynomial time)-complete problems — those for which a potential solution can be verified in a reasonable amount of computer time; but the problem can cost exponential time to solve in the worst case. Thousands of computational problems from many different fields can be shown to be instances of the K-SAT problem. In fact, K-SAT problems are often used to benchmark the performance of search algorithms. These problems can always be put in a canonical form. There are usually N logical variables that may be given the values 'true' and 'false', and a set of M clauses, which are 'or' statements about the N variables. If the problem is pure K-SAT then these clauses connect K of the variables. For instance, in my little example above, the fact that there are three names translates into a three-'or' statement that the name can be Alice or Barbara or Carol. Then for the pure K-SAT problem one asks that the logical sum of all the clauses be true. Monasson *et al.*¹ also consider the problem of determining the minimum number of clauses that need to be falsified — crossed out — to make the sum of the clauses true.

This work is the latest of many attempts^{2–4} to forge a connection between the classic problems of computer complexity theory and the statistical mechanics of random systems, such as the 'spin glasses' — a type of alloy containing a random distribution of magnetic atoms, the spins of which are 'frustrated', which is to say they have conflicting interactions.

In the case of K-SAT, the corresponding spin-glass energy function can simply be constructed by making the variables into spins — 'up' equals 'true', and so on — and by counting the number of unsatisfied clauses for any configuration of the spins. Because sophisticated mathematical methods have evolved for solving such statistical-mechanics problems, one might hope that this would have important implications for complexity theory. Unfortunately, these hopes have been mostly unsatisfied, because the goals of the two 'solutions' are so different. The computer scientist's solution is a particular configuration in a particular instance, and his proofs must encompass all difficult cases; the physicist is looking at a typical case among some universe of randomly chosen examples, and at the statistically average result for large N .

Nonetheless, some ideas from statistical mechanics have been helpful. Notably the 'simulated annealing' algorithm has been very effective in tackling the well-known 'travelling salesman problem', another NP-complete problem, which has many commercially important applications. This algorithm is based on the physics of phase transitions, specifically the way that the structure of a glassy solid relaxes as it is cooled. To reduce internal stresses the trick is to slowly cool the glass through the transition (the process of annealing), so that it is able to find

its minimum energy state. If the glass is cooled too quickly it ends up in a higher energy state. Equivalently, if a search algorithm approaches a solution too rapidly, it may find a local, but not necessarily a global minimum.

The general idea of phase transitions turns out to apply to K-SAT as well, in that there is a critical value of α , defined by $\alpha = M/N$, for each value of K , below which almost all cases are satisfiable and above which they are almost all unsatisfiable. In this sense there is a discontinuous 'phase transition' across which drastic changes occur in the computational difficulty. The value of α for $K = 2$, as well as a number of other average properties, may be found analytically. It is an interesting and important fact that it is at values of α near the transition that the average cost of heuristic searches for the solution is maximum, making it extremely hard to find a solution³. But at values of α away from the boundary it becomes relatively easy to solve the formulae (by showing them to be satisfiable and unsatisfiable at low and high values respectively).

Monasson *et al.*¹ concentrate on '2 + p - SAT' problems: where a fraction, $p < 1$, of the clauses have three variables, the rest two. The basic 2-SAT problem is not NP-complete, but 'polynomial' and therefore easy to solve, whereas 3-SAT is NP-complete. All problems with $p > 0$ are NP-complete according to mathematicians, but by empirically testing a large sample of randomly generated cases Monasson *et al.* find that, below $p = 0.41$, the computational cost scales linearly with N , and a critical value of α corresponding to a continuous transition from satisfiable to unsatisfiable solutions can be

Developmental biology

Bonsai flies

Miniature organisms are no longer the sole preserve of Japanese bonsai lovers. Ernst Hafen and colleagues (*Cell* 97, 865–875; 1999) have discovered that flies with mutations in a gene termed *chico* — meaning 'small boy' in Spanish — are less than half the normal size (compare the *chico* mutant fly with the wild-type fly below it). Not only are the cells in *chico*[−] flies smaller than usual, but there are fewer of them — defects that the authors attribute to decreased cell growth and proliferation rather than increased cell death.

The Chico protein is, in fact, very similar to the vertebrate IRS1–4 proteins, which are involved in cellular metabolism. Consistent with this, Hafen's group find that *chico*[−] flies also have metabolic defects — despite their diminutive stature, they contain twice as much lipid as normal. It

also turns out that *chico*[−] flies are similar in size to normal flies reared without proper nutrition. So, the authors propose that Chico may belong to a nutritional sensing system which, by regulating cell proliferation, growth and metabolism, controls growth in response to nutritional conditions.

Alison Mitchell



found analytically. But at $p = 0.41$ the phase transition changes character and simultaneously the computational cost of heuristic searches becomes exponentially great. The authors have defined an order parameter for the statistical phase transition, and above $p = 0.41$ they show that this phase transition jumps discontinuously at the transition. But its variation near the transition suggests that there must be critical fluctuations. (The incorrect statement is made¹ that this is new, but in fact it resembles markedly and unexpectedly the Anderson–Yuval–Kosterlitz–Thouless type of defect-mediated transition⁵, used to describe melting in two dimensions.)

In addition to the fascinating implications of this unexpected structure for computational complexity theory, where it marks a long-suspected distinction between

problems that are in some sense ‘usually’ simple to solve rather than ‘usually’ difficult, this is the first direct correlation of a transition in computational complexity with a critical point of a statistical-mechanics transition. For statistical physicists, this ‘new’ type of transition (but see my parenthetical comment above) may hold unexpected implications. □

Philip W. Anderson is in the Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544-0708, USA.

e-mail: pwa@pupgg.princeton.edu

1. Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B. & Troyansky, L. *Nature* **400**, 133–137 (1999).
2. Mezard, M., Parisi, G. & Virasoro, M. A. *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
3. Kirkpatrick, S. & Selman, B. *Science* **264**, 1297–1301 (1994).
4. Monasson, R. & Zecchina, R. *Phys. Rev. Lett.* **76**, 3881–3885 (1996).
5. Thouless, D. J. in *Topological Quantum Numbers in Non-Relativistic Physics* 109–115 (World Scientific, Singapore, 1998).

Alzheimer's disease

Antibody clears senile plaques

Peter H. St George-Hyslop and David A. Westaway

People suffering from Alzheimer's disease develop a progressive dementia in adulthood, accompanied by three main structural changes in the brain: diffuse loss of neurons in the hippocampus and neocortex; accumulation of intracellular protein deposits termed neurofibrillary tangles; and accumulation of extracellular protein deposits termed amyloid or senile plaques, surrounded by misshapen nerve terminals (dystrophic neurites). A main constituent of these amyloid plaques is the amyloid- β peptide ($A\beta$), a 40–42-amino-acid protein that is produced through cleavage of the β -amyloid precursor protein (APP). Although Alzheimer's disease can be treated, we can currently neither prevent nor cure it. On page 173 of this issue, however, Schenk *et al.*¹ show that, in a mouse model of Alzheimer's disease, immunization with $A\beta$ inhibits the formation of amyloid plaques and the associated dystrophic neurites.

These results raise the possibility of vaccination with $A\beta$ against human Alzheimer's disease. But before this can be seriously entertained, several questions must be answered. Schenk and colleagues¹ found that high levels of anti-human $A\beta$ antibody were necessary for the effect to be seen in mice. So, can injection with human $A\beta$ induce enough of the antibody? Will immune tolerance (whereby the immune system does not react against the body's own proteins) frustrate this, as it has often done with attempts to target cancers with antibodies? Conversely, is it safe to immunize people with high levels of a protein that is widely expressed outside the protective confines of the blood–brain barrier? When Schenk *et al.* immunized the mutant mice (known as PDAPP mice, because they over-

express a human APP transgene bearing the pathogenic valine-to-phenylalanine mutation at position 717) with human $A\beta$ peptide, they did not see any autoimmune responses. But then, the human $A\beta$ antigen induced much lower levels of antibodies to the endogenous mouse $A\beta$ or APP.

The most critical question is whether depletion of the amyloid plaques is accompanied by an improvement in the behavioural/neurophysiological impairments,

and a reduction in the nerve cell death of Alzheimer's disease? In other words, does immunization with $A\beta$ simply clear a neuro-pathological by-product or can it cure the disease? This question may be difficult to answer because all of the current animal models (based on overexpression of human APP and/or presenilin-1 transgenes bearing missense mutations associated with Alzheimer's disease) provide only a partial model of the human condition. So, although these animals accumulate increased levels of $A\beta$ in the brain and have many amyloid-plaque deposits, they have only subtle behavioural and electrophysiological deficits. More problematically, these animals do not develop neurofibrillary tangles or show significant neurodegeneration (refs 2–4 and D. A. Westaway *et al.*, unpublished observations).

A second set of questions concerns the mechanism by which immunization with $A\beta$ blocks the formation of amyloid plaques. Antibodies against $A\beta$ might act as an artificial chaperone for extracellular $A\beta$, possibly by binding to $A\beta$ and preventing it from aggregating or from changing into β -pleated-sheet conformation. Alternatively, these antibodies could accelerate clearance of $A\beta$ from the central nervous system through one of several peripheral mechanisms (targeting $A\beta$ for destruction by the peripheral reticuloendothelial system, for example, or reducing the production of $A\beta$ in the periphery). Finally, $A\beta$ might affect immune modulation of inflammatory mechanisms that are thought to be activated in Alzheimer's disease.

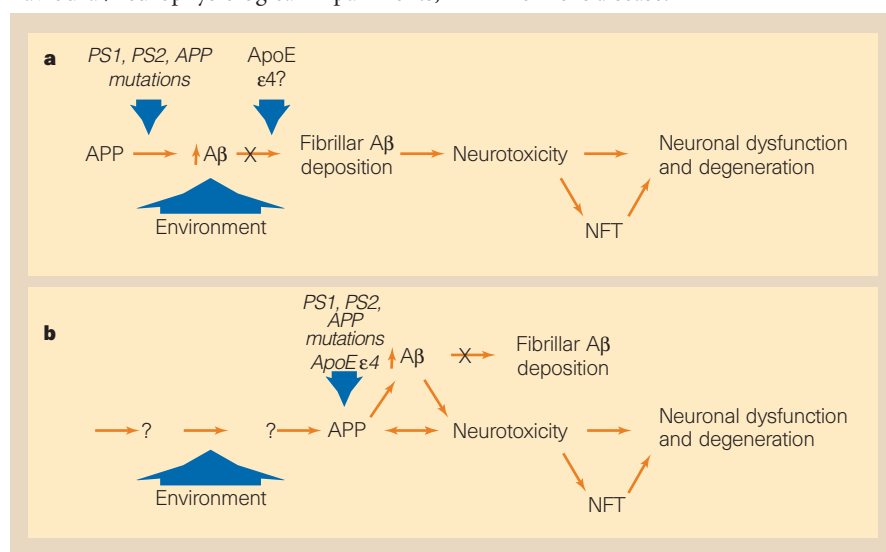


Figure 1 Controversies about the neuropathological alterations in Alzheimer's disease. a, In the 'amyloid cascade' hypothesis, generation of extracellular amyloid- β peptide ($A\beta$) from the β -amyloid precursor protein (APP) is thought to be central to the pathogenesis of Alzheimer's disease. b, An alternative point of view is that the extracellular plaques containing $A\beta$ are an invariant but peripheral event. Large arrows depict putative entry points of environmental and genetic causes of Alzheimer's disease in these pathways. Genetic causes include mutation or polymorphisms in presenilin-1 (PS1), presenilin-2 (PS2), APP and Apolipoprotein E (ApoE). Schenk *et al.*¹ have found that, in a mouse model of Alzheimer's disease, immunization with $A\beta$ presumably blocks (marked with an X) the extracellular deposition of this protein. (NFT, neurofibrillary tangles.)

Addressing these questions will almost certainly shed new light on an old debate about the pathogenesis of Alzheimer's disease. It is generally accepted that neuronal dysfunction and degeneration underlie the clinical dementia of the disease. But there has long been a controversy as to whether extracellular deposition of fibrillar A β in amyloid plaques is part of the biochemical processes that cause the neuronal dysfunction and death, or whether it is merely a by-product of this process.

One school of thought (the 'amyloid cascade' hypothesis) maintains that the processes that cause accumulation of A β are central to the pathogenesis of Alzheimer's disease (Fig. 1a). It is unclear whether extracellular accumulation of A β (and especially of the longer, potentially more toxic, A β_{42} species) is more important than intracellular accumulation. Nevertheless, this hypothesis is supported by several observations. First, in some inherited forms of Alzheimer's disease, missense mutations in APP, presenilin-1 and presenilin-2 (refs 5–7) lead to overproduction of A β (ref. 8). Second, A β_{42} is toxic both when injected into mammalian brain (especially fibrillar A β_{42} in aged primate brain⁹) and when administered to cultured neurons (reviewed in ref. 10). Finally, deposition of A β is the earliest neuropathological change detected in people who have mutations in presenilin-1 but do not yet show symptoms of Alzheimer's disease¹¹.

Opponents of this hypothesis argue that although deposition of extracellular A β is an almost invariant event in the pathogenesis of Alzheimer's disease, it need not be the prime cause of the dementia (Fig. 1b). They argue first that the density of the A β plaques correlates only poorly with severity of the dementia^{12,13}. Second, the degree of dementia is better correlated with the density of cerebral neurofibrillary tangles¹³. Finally they state that, in some of the APP- and presenilin-1-based transgenic mouse models of Alzheimer's disease, the subtle functional deficits occur before the formation of amyloid plaques (ref. 4 and N. Agopyan *et al.*, unpublished observations). Moreover, some cases of Alzheimer's disease lack the typical amyloid plaques¹⁴.

The ability to modulate deposition of extracellular A β in the central nervous system while monitoring behavioural and electrophysiological changes may now allow us to determine whether extracellular A β deposits are the real villain in Alzheimer's disease. Moreover, if immunization with A β is indeed useful against Alzheimer's disease, similar strategies might be applied in the other diseases where extracellular plaques are formed. These might include some forms of prion disease (although, as with Alzheimer's disease, some experiments indicate that extracellular plaques are not neurotoxic¹⁵), and amyloidotic polyneuropathies

(where transthyretin accumulates). This knowledge could even be useful in treating conditions where intracellular deposits are formed, such as Parkinson's disease (α -synuclein) and frontotemporal dementia (tau). □
Peter H. St George-Hyslop and David A. Westaway are at the Centre for Research in Neurodegenerative Diseases, Departments of Medicine (Neurology) and Laboratory Medicine and Pathobiology, University of Toronto, 6 Queen's Park Crescent West, Tanz Neuroscience Building, Toronto, Ontario M5S 3H2, Canada.

e-mail: p.hyslop@utoronto.ca

Condensed-matter physics

Simple metals under pressure

Richard M. Martin

The alkali metals — the monovalent elements in the first column of the periodic table — have long been regarded as prototype 'simple metals' with non-directional metallic bonding. This view is entrenched in the current understanding of solids, which began with the pioneering work of Wigner and Seitz¹. Under ordinary conditions, all the alkalis from lithium to caesium occur in simple body-centred-cubic (bcc) or close-packed lattices, typical of 'nearly-free electron' metals. The only exception is hydrogen, where protons pair to form H₂ molecules and a condensed molecular solid that is insulating. Because of its place as the first element in the periodic table, it is a Holy Grail of physics to compress hydrogen by pressure until it becomes metallic, like the other alkali metals. It was anticipated that under pressure hydrogen would adopt simpler structures and become closer packed; but such metallic phases have not yet been found by experiments, which have instead revealed transitions to low-symmetry molecular structures². On page 141 of this issue, Neaton and Ashcroft³ report the unexpected finding that under pressure, lithium (Li) instead acts like H₂ — that is, the nuclei are predicted to form pairs producing structures similar to condensed phases of H₂.

The predictions of Neaton and Ashcroft are based on state-of-the-art electronic structure calculations, which show that the bcc phase of Li should transform to an orthorhombic structure near 50 gigapascals (GPa) and to a molecular, semimetallic phase near 100 GPa, finally transforming back to a monatomic metal again at very high pressures. The molecular-like structures favoured by Li at intermediate pressures are related to those thought to occur in molecular hydrogen near its insulator-metal transition. Preliminary calculations³ also indicate that similar transitions should occur in sodium at much higher pressure and with a much smaller degree of molecular pairing. If this behaviour is actually confirmed by experiments, it will be

1. Schenk, D. *et al.* *Nature* **400**, 173–177 (1999).
2. Masliah, E. *et al.* *J. Neurosci.* **16**, 5795–5811 (1996).
3. Chapman, P. F. *et al.* *Nature Neurosci.* **2**, 271–276 (1999).
4. Moechars, D. *et al.* *J. Biol. Chem.* **274**, 6483–6492 (1999).
5. Goate, A. M. *et al.* *Nature* **349**, 704–706 (1991).
6. Sherrington, R. *et al.* *Nature* **375**, 754–760 (1995).
7. Rogae, E. I. *et al.* *Nature* **376**, 775–778 (1995).
8. Scheuner, D. *et al.* *Nature Med.* **2**, 864–870 (1996).
9. Guelin, C. *et al.* *Nature Med.* **4**, 827–831 (1998).
10. Yankner, B. A. *Neuron* **16**, 921–932 (1996).
11. Lipka, C. F. *et al.* *Lancet* **352**, 1117–1118 (1998).
12. Berg, L. *et al.* *Arch. Neurol.* **55**, 326–335 (1998).
13. Terry, R. D., Masliah, E. & Hansen, L. A. in *Alzheimer Disease* (eds Terry, R. D., Katzman, R. & Bick, K. L.) 179–196 (Raven, New York, 1994).
14. Crook, R. *et al.* *Nature Med.* **4**, 452–455 (1998).
15. Brandner, S. *et al.* *Nature* **379**, 339–343 (1996).

an achievement for modern theoretical methods, leading to a revised understanding of the alkalis: the apparently simple picture under ordinary conditions being the anomaly, and the range of behaviour under pressure providing new challenges and opportunities.

The 'nearly-free electron' behaviour of the valence electrons of the alkalis is revealed most clearly by their Fermi surfaces, which describe the boundary between occupied and empty electronic states, and which very nearly resemble the sphere that would occur if there were no influence whatsoever from the ion cores. This behaviour is understood in terms of weak electron-ion interactions or 'pseudopotentials' that describe the effective interaction of valence electrons with the ion cores. However, there have been many signs that the alkalis are not so simple. One mystery is the apparent absence of superconductivity, despite the fact that the interactions of the electrons with the lattice vibrations appear to be strong enough to lead to conventional superconductivity, especially in Li (ref. 4). In order to explain such anomalies, large modifications of the Fermi surface have been proposed⁵ that would be manifested as unusual optical properties and suppression of superconductivity. It has long been known that deviations of the Fermi surface of Li require a non-local pseudopotential that acts differently for electronic states with zero angular momentum (that is, *s*-states), as opposed to states with higher angular momentum (such as *p*- or *d*-states).

The electronic structure of Li under extreme pressure was calculated by Boettger and Trickey years ago⁶. They found exceptionally large deviations of the Fermi surface from a sphere (also seen by Neaton and Ashcroft; see their Fig. 4 on page 143), which they interpreted as an '*s*-*p*' transition — in which the lowest-energy electronic states change from primarily *s*-like to primarily *p*-like, producing an effective change in the chemistry of Li. Promotion of electrons to unusual valence configurations was observed long ago in another

alkali metal, caesium. The classic experiments of Bridgman in the 1940s revealed an anomalous volume discontinuity under pressure. As early as 1950, Sternheimer⁷ interpreted this as an 's-d' transition in which the electronic d-states became lower in energy relative to the s-states under pressure (see also refs 8,9). Sternheimer attributed the general idea to Fermi (without a citation) saying, "In order to explain the phase transition, Fermi proposed that the valence electron is forced into a vacant internal orbit".

Neaton and Ashcroft³ find that changes in electronic structure also lead to remarkable changes in the crystal structure, including distorted and paired hydrogen-like structures that may even be insulating. Up to now theorists have considered only high-symmetry 'simple' structures; within the confines of such structures the only phase transitions are between bcc and close-packed structures that must have simple metallic behaviour.

Neaton and Ashcroft have offered an alternative interpretation for the transition in Li under pressure. They propose that in any element the valence electrons are excluded from the ion core regions by the Pauli exclusion principle, which prevents any two electrons from occupying the same state. This effect grows with pressure as the volume available to the valence electrons decreases, and Neaton and Ashcroft propose that this effect is especially pronounced in Li because of its large core (the size of the core decreases as the nuclear charge increases). This also leads to a strong repulsive pseudopotential that can cause a distortion of the lattice. In high-symmetry structures there will be multiple electronic states with the same energy that are split by the distortion, with the electrons occupying only the lowest-energy state. This is the solid-state analogue of the Jahn-Teller effect in molecules, which can always lower their energy by such a distortion. As shown in their Fig. 5 on page 143, Neaton and Ashcroft find that in the molecular-like state the valence electrons are excluded from the region between the pairs of nuclei and instead occupy states with maximum density in the interstitial regions. This picture could be viewed as a different interpretation of Fermi's idea⁷ that under pressure the valence electron is forced into a vacant internal orbit.

There are many experimental consequences of the calculations of Neaton and Ashcroft and experimental tests are already underway¹⁰. Among the predicted observable effects are large changes in the optical behaviour of Li from a silvery metallic appearance to a transparent or black colour, as expected for an insulator or semimetal. Preliminary measurements to be announced this month by V. V. Struzhkin at the *International Conference on High Pressure Science and Technology* in Honolulu, Hawaii, suggest new evidence for such effects. If the structures have low sym-

metry like those in molecular hydrogen, then there may also be strong infrared absorption as has been discovered in hydrogen².

Part of the excitement over this work is that if Li does form molecular analogues of hydrogen under pressure, then there is the possibility of superconductivity at a relatively high temperature for an element, and other interesting phenomena long sought for hydrogen. Because the effects occur at a much lower pressure than in hydrogen, perhaps Li can point the way to creating the Holy Grail of metallic hydrogen under pressure. Finally, such behaviour will clearly move Li and other alkalis away from being 'simple metals' in the

minds of condensed-matter physicists and towards the category of 'interesting metals'.

Richard M. Martin is in the Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA.
e-mail: rmartin@uiuc.edu

1. Wigner, E. P. & Seitz, F. *Phys. Rev.* **43**, 804-810 (1933).
2. Mao, H. K. & Hemley, R. J. *Rev. Mod. Phys.* **66**, 671-692 (1994).
3. Neaton, J. B. & Ashcroft, N. W. *Nature* **400**, 141-144 (1999).
4. Lang, K. M. *et al.* *J. Low Temp. Phys.* **114**, 445-454 (1999).
5. Overhauser, A. W. *Phys. Rev. Lett.* **53**, 64-65 (1984).
6. Boettger, J. C. & Trickey, S. B. *Phys. Rev. B* **32**, 3391-3398 (1985).
7. Sternheimer, R. *Phys. Rev.* **78**, 235-243 (1950).
8. Louie, S. G. & Cohen, M. L. *Phys. Rev. B* **10**, 3237-3245 (1974).
9. McMahan, A. K. *Phys. Rev. B* **17**, 1521-1527 (1978).
10. Struzhkin, V. V., Hemley, R. J. & Mao, H. K.; Mori, Y. & Ruoff, A. L. *Bull. Am. Phys. Soc.* **41**, 1489 (1999).

Chloroplasts

Ever decreasing circles

Geoff McFadden

If there is life on Mars, it may be disappointingly ordinary compared to some bizarre earthlings. Consider the genetic blueprint of dinoflagellates. The nucleus of these marine algae has DNA with non-standard bases, lacks typical histones (DNA packaging proteins common to all other nucleated cells) and has permanently condensed DNA, as if perpetually poised for division. But if the dinoflagellate nucleus is peculiar then, as we find on page 155 of this issue¹, the chloroplast is even more so. There Zhang *et al.* present the first analysis of the structure and gene content of the dinoflagellate chloroplast genome, one of the last frontiers in understanding chloroplast evolution.

Chloroplasts constitute the photosynthetic machinery of eukaryotes. They derive from endosymbiotic cyanobacteria and bear hallmarks of that prokaryotic ancestry². The chloroplast genome is smaller than that of a cyanobacterium, but is still circular with a single replication origin and genes assembled into operons (expression units incorporating numerous, often related, genes). Until now, all known chloroplast genomes adhered to this uniform architecture, but dinoflagellates buck the trend.

Zhang *et al.*¹ show that, despite a common origin with other chloroplasts, dinoflagellate chloroplasts have adopted a unique genome architecture in which each gene has its own minicircle. In other words, dinoflagellate chloroplasts contain a small family of plasmid-like molecules, each carrying a gene for a protein or ribosomal RNA. The authors¹ find that all minicircles in a given dinoflagellate species have a repeat region that appears to participate in inter-circle gene conversion, thereby maintaining identity between circles, and they propose that this region contains a replication origin and an anchor point for circle segregation during chloroplast division. This repeat region could be a key to development of a genetic

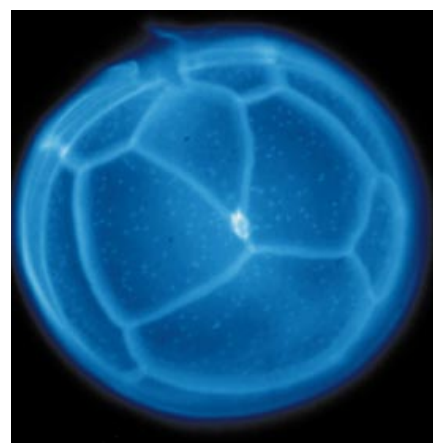


Figure 1 The dinoflagellate *Diplopelta bomba*. The chloroplast genome of dinoflagellates is like no other, and long defied characterization, but its genetic blueprint has now been revealed by Zhang *et al.*¹.

transformation system for dinoflagellate chloroplasts — it should be easy to introduce a plasmid-type construct, incorporating the repeat and a selectable marker, into them.

How dinoflagellates developed minicircles from a large, multigenic circle (the ancestral state for chloroplasts) is not apparent. But there is an intriguing similarity in another endosymbiotic organelle, the mitochondrion. Mitochondrial genomes typically resemble chloroplast genomes (and their respective bacterial progenitor genomes) in being circular³. However Watanabe *et al.*⁴ find that the single-gene minicircles in mitochondria of a little-known group of miniature parasitic animals, known as mesozoa, do not share a common non-coding region, although the mitochondrial circles do possess stem loop structures that may function in replication/segregation. There is no evidence for a canonical master circle in either the dinoflagellate chloroplast or the meso-

GERT HANSEN

zoan mitochondrion, but neither group^{1,4} can rule out its existence.

Gene content of the dinoflagellate chloroplast seems minimal. Normal chloroplasts retain some 100–200 genes², but Zhang *et al.*¹ found only nine genes among all the minicircles they examined. Other, less abundant circles probably await discovery, but we can expect the gene catalogue to be impoverished. Where are the missing genes? In the nucleus, no doubt. A pervasive trend in endosymbiosis is confiscation of the chloroplast's genes by the nucleus (the host), which is estimated to hold 800–900 plastid protein genes². Why then has the dinoflagellate nucleus asserted even more control over its photosynthetic slave than other hosts? Minicircles may be the key. Transfer of DNA may have been expedited by each gene being packaged on a discrete, compact unit better able to make the journey from one part of the cell to another. In dinoflagellates, the products of these vagrants must be copious, so tracing them might be as simple as randomly sequencing active nuclear genes (an expressed sequence tag approach).

A revelation of molecular phylogeny was that dinoflagellates are close relatives of human parasites such as *Plasmodium* (which causes malaria) and *Toxoplasma*⁵ — and here the plot really thickens. These parasites have a relict chloroplast^{6–8}, so could the chloroplasts in *Plasmodium* and dinoflagellates have the same origin? Zhang *et al.*¹ provide the data from dinoflagellates to help answer that question, but all is not yet clear. Ironically, the *Plasmodium* chloroplast genome (which is circular and encodes 68 genes⁶) is more conventional than that of dinoflagellates, preventing whole-genome comparison. Scrutiny of individual chloroplast genes shared by *Plasmodium* and dinoflagellates can reveal little more. *Plasmodium* chloroplast genes are highly divergent⁶, and the dinoflagellate chloroplast genes are even more so. Because divergent genes tend to be grouped artificially in the calculations involved in building phylogenetic trees⁹, any grouping of *Plasmodium* and dinoflagellate chloroplast genes must be treated with scepticism.

So we still cannot tell if dinoflagellates and *Plasmodium* have the same chloroplast. But we do have further insight into the origin of dinoflagellate chloroplasts, which are suspected to have been acquired by a process known as secondary endosymbiosis⁹. Zhang *et al.* provide strong supporting evidence for that view. The endosymbiosis of a cyanobacterial-like cell within a eukaryote to create the original chloroplast is referred to as the primary endosymbiosis (and in another paper in this issue, on page 159, Tomitani *et al.*¹⁰ provide compelling evidence that a single primary endosymbiosis is ultimately the source of all chloroplasts). Secondary endosymbiosis is the subsequent purloining of chloroplasts by non-photosynthetic

eukaryotes that engulf and retain a (primary) chloroplast-containing cell; the process occurred frequently in eukaryotic evolution and leaves a tell-tale clue in the form of multiple chloroplast membranes⁹.

By analysing chloroplast genes, Zhang *et al.* show that dinoflagellates, whose plastids have three membranes, probably engulfed a red-algal-like cell. An independent study comes to the same conclusion¹¹. Nonetheless these exciting results do not solve the origin of the *Plasmodium* chloroplast, which has four membranes and was also acquired secondarily^{8,12}, and there is vigorous debate over whether it derived from a red alga¹³ or a green alga⁸. This issue is of more than academic interest because the *Plasmodium* chloroplast could be an ideal target for drug therapies. Many drugs that inhibit chloroplast activities kill *Plasmodium* and *Toxoplasma*¹⁴,

so increased understanding of chloroplasts could ultimately help combat malaria and related infections. □

Geoff McFadden is in the Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville 3052, Australia.
e-mail: g.mcfadden@botany.unimelb.edu.au

1. Zhang, Z., Green, B. R. & Cavalier-Smith, T. *Nature* **400**, 155–159 (1999).
2. Martin, W. *et al.* *Nature* **393**, 162–165 (1998).
3. Gray, M., Burger, G. & Lang, B. *Science* **283**, 1476–1481 (1999).
4. Watanabe, K. *et al.* *J. Mol. Biol.* **286**, 645–650 (1999).
5. Wolters, J. *Biosystems* **25**, 75–84 (1991).
6. Wilson, R. J. M. *et al.* *J. Mol. Biol.* **261**, 155–172 (1996).
7. McFadden, G. I. *et al.* *Nature* **381**, 482 (1996).
8. Köhler, S. *et al.* *Science* **275**, 1485–1488 (1997).
9. Palmer, J. D. & Delwiche, C. F. *Proc. Natl Acad. Sci. USA* **93**, 7432–7435 (1996).
10. Tomitani, A. *et al.* *Nature* **400**, 159–162 (1999).
11. Takishita, K. & Uchida, A. *Phycol. Res.* (in the press).
12. Waller, R. F. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 12352–12357 (1998).
13. Blanchard, J. J. *Euk. Microbiol.* **46**, 367–375 (1999).
14. McFadden, G. I. & Roos, D. S. *Trends Microbiol.* (in the press).

Evolutionary biology

Dirty eating for healthy living

Jared M. Diamond

As babies, we are warned by our mothers not to eat dirt, but as adults some of us do it anyway and dignify it with the name of geophagy. The regular and intentional consumption of soil, by itself or mixed with food, has been recorded from traditional human societies on all continents, especially among pregnant women^{1–4}. Geophagy has also been documented in many species of mammals, birds, reptiles, butterflies and isopods, especially among herbivores^{5–9}. Why do they and we do it? Proposed biological functions of geophagy have now been tested by James Gilardi and co-workers¹⁰, who uncover a fascinating evolutionary arms race between plants and their would-be animal consumers.

The dirt-eaters studied were Peruvian Amazon rainforest parrots, of which a thousand or more individuals of 21 species gather early each morning at certain sites with exposed bare soil on river banks or cliff faces (Fig. 1). Because these sites are ideal for viewing and photography, they attract 4,000 bird-watching tourists each year, support 500 jobs in the local ecotourism industry, and earn Peru about US\$1,000 per year per individual wild macaw. The birds' taste in dirt is highly specific: for instance, they congregate not just at one particular bend of the Manu River but at one soil band running hundreds of metres horizontally along that bend, spurning the dirt in bands one metre above or below the preferred band. Gilardi *et al.* tested possible functions of geophagy by comparing the physical and chemical properties of soil samples from the preferred and rejected bands.

The commonest explanation for geophagy in birds is to provide grit⁸. Because birds lack teeth, many ingest pebbles or

coarse soil with which to grind food in their gizzards. Preferred particle sizes of grit increase with bird size, from 0.5 mm for sparrows to 2.5 cm for ostriches. However, Gilardi *et al.* found that the soil preferred by Peruvian parrots is very fine: only 5% of it by volume is coarse sand exceeding even 0.05 mm in particle diameter. Most of it is clay less than 0.2 µm in particle diameter, and preferred soils contain only a quarter as much coarse sand and nearly twice as much fine clay as rejected soils. So parrots are not eating soil to get grit. On reflection, this is not surprising: parrots have no need for grit because their strong, sharp bills can shred the hardest nuts.

A second function of geophagy, suggested for livestock, wild ungulates, rabbits, butterflies and pregnant women, is to provide essential minerals^{6,7}. Soils sold in Ghanaian markets to pregnant African women are richer in iron and copper than the dietary supplement pills made by pharmaceutical companies specifically for prenatal use. But Gilardi *et al.*¹⁰ found that soils preferred by parrots contain lower available quantities of most biologically significant minerals than non-preferred soils, and much lower quantities than the parrots' preferred plant foods. Hence, unless the parrots are making a big mistake in their taste preferences, they are not selecting soils for mineral content.

A third function of geophagy, proposed for ungulate livestock, is to buffer the rumen contents⁶. Because parrots lack a rumen, it will come as no surprise that their preferred soils have no more buffering capacity than distilled water.

What, then, do the parrots actually gain from ingested soil? It turns out that they regularly eat seeds and unripe fruits whose con-



Figure 1 Blue-headed parrots (*Pionus menstruus*) at a clay lick in Manu, Peru. Gilardi *et al.*¹⁰ have shown that minerals in the clay detoxify the birds' plant diet.

tent of alkaloids and other toxins renders them bitter and even lethal to humans and other animals. Because many of these chemicals are positively charged in the acidic conditions found in the stomach, they bind to clay minerals bearing negatively charged cation-exchange sites^{2,3,5,9}. That's why experienced tourists visiting destinations with poor sanitation carry medicines such as kaopectate (high in clay minerals) to adsorb the toxins. That's also why peasant farmers and hunter-gatherers throughout the world often mix bitter but otherwise nutritious plant foods (like acorns and wild potatoes) with selected soils before consumption¹⁻³.

Peruvian parrots behave like sophisticated human tourists and hunter-gatherers. Their preferred soils were found to have a much higher cation-exchange capacity than adjacent bands of rejected soils — because they are rich in the minerals smectite, kaolin and mica. In their capacity to bind quinine and tannic acid, the preferred soils surpass the pure mineral kaolinite and surpass or approach pure bentonite. Clearly, parrots would be well qualified for jobs as mining prospectors.

Gilardi *et al.* confirmed this hypothesis with two sets of bioassays. First, they exposed brine shrimp (the toxicologist's test animal of choice) to extracts of seeds routinely consumed by macaws. Many of the brine shrimp died, confirming the toxicity of the parrots' diet. But mixing the solutions or extracts with soil preferred by parrots reduced the effective toxin loads by 60–70% and improved shrimp survival. Second, Amazon parrots were given an oral dose of the alkaloid quinidine with or without preferred soil, and quinidine levels were measured in the parrots' blood for three hours as absorption

took place from the gut. Providing soil along with the quinidine reduced absorbed quinidine blood levels by 60%.

What is the evolutionary significance of plant toxins and animal anti-toxin behaviour? From a plant's evolutionary perspective, a seed should be high in nutrients to support germination and seedling growth; the ripe fruit around the seed should also be nutrient-rich and attractive to animals, encouraging them to pluck and eat the fruit and disperse the seed. On the other hand, the seed itself should be repulsive to animal consumers, inducing them to regurgitate or defaecate it, and the unripe fruit should be repulsive, lest animals harvest it before the seed is viable. From an animal's evolutionary perspective, an ability to defeat the plant's toxin defences would enable it to obtain the nutrients in the seed as well as those in the ripe fruit, and to outcompete other animal consumers by harvesting the fruit while it is unripe and still unpalatable to them.

Any textbook of animal biology describes the resulting evolutionary arms race, in which plants evolve increasingly potent toxins (such as strychnine and quinine), and animals evolve increasingly potent means of detoxification. While enzymatic detoxification has previously received the most attention, the work of Gilardi *et al.*¹⁰ and the wide distribution of geophagy among animal herbivores suggest an additional important means of detoxification by adsorption on ingested soil minerals.

A host of interesting questions now comes into focus. How do parrots discover the best soils — can they discriminate among soils immediately by texture and taste, or must they experiment with various soils mixed with toxic food and discover which soil assuages their upset stomach? Might the availability of suitable geophagy sites limit herbivore distributions and merit concern from conservation biologists? Only certain species of local herbivores are reported as visiting geophagy sites: why? To return to our youthful dirty habits, do curious dirt-licking babies deserve our encouragement for their experiments with self-medication?

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA. e-mail: jdiamond@mednet.ucla.edu.

1. Johns, T. J. *Chem. Ecol.* **12**, 635–646 (1986).
2. Johns, T. *With Bitter Herbs They Shall Eat It: Chemical Ecology and the Origins of Human Diet and Medicine* (Univ. Arizona Press, Tucson, 1990).
3. Johns, T. & Duquette, M. *Ecol. Food Nutr.* **25**, 221–228 (1991).
4. Abrahams, P. W. & Parsons, J. A. *Geogr. J.* **162**, 63–72 (1996).
5. Hladik, C. M. & Gueguen, L. C. *R. Acad. Sci.* **279**, 1393–1396 (1974).
6. Jones, R. L. & Hanson, H. C. *Mineral Licks, Geophagy, and Biogeochemistry of North American Ungulates* (Iowa State Univ. Press, Ames, 1985).
7. Kreulen, D. A. *Mamm. Rev.* **15**, 107–123 (1985).
8. Gionfriddo, J. P. & Best, L. B. *Condor* **97**, 57–67 (1995).
9. Mahaney, W. C. *et al.* *Primates* **37**, 121–134 (1996).
10. Gilardi, J., Duffey, S., Munn, C. & Tell, L. *J. Chem. Ecol.* **25**, 897–922 (1999).

Daedalus

Creep and anti-creep

Many substances — ice is perhaps the best known — can crystallize in lots of different ways, many of them stable only under high positive pressure. Daedalus is now exploring the converse field, that of crystals stable only under negative pressure, that is, tension. Engineering components are routinely stressed to thousands of atmospheres of tension: usually in one or two dimensions only, but isotropic three-dimensional tension is possible. Even liquids, if clean and degassed, can be tensioned to many hundreds of negative atmospheres. Indeed, the sap in trees taller than ten metres is thought to be under permanent tension.

So DREADCO physicists are melting numerous solids, putting the liquid under strong tension, and letting them resolidify again. They are also submitting crystalline samples to sustained three-dimensional tension, and looking for a slow phase-change to some expanded, negative-pressure crystal habit. The pilot experiments are largely empirical; it is hard to guess which substances form distinctive negative-pressure phases. But Daedalus hopes that at least some of these phases will continue to exist metastably at atmospheric pressure, at least for a while.

His ultimate goal is a new engineering material. Many such materials, he points out, creep under load. For a component in tension this is a dangerous vice. But for one in compression it can be a virtue. If overloaded, it creeps plastically away from the load, thickening as it does so, and sharing its burden with more lightly loaded members nearby. A compression structure is often usefully 'self-designing'.

Metastable expanded materials should bring the same self-optimization to tension structures. While it remains tensioned, a component of such a material will be quite stable. But if the tension slackens, it will become metastable. It will slowly contract to its denser phase, restoring the tension and relieving nearby members of some of their load. In fact it will show 'anti-creep'.

DREADCO's anti-creep alloys will be widely welcomed. Bridges, bicycles, power lines, aerospace frames, all will exploit anti-creep technology for greater safety and efficiency. Self-tightening anti-creep fasteners and connectors will transform the small-scale details of engineering. Over the whole field, designers will gratefully allow self-optimizing anti-creep materials to lift some of their lonely burden.

David Jones

Obituary

Oleg Ptitsyn (1929–99)

Pioneer in protein folding

Oleg Ptitsyn was one of those rare scientists who are prepared to advance new but unfashionable ideas, and pursue them until they either fall by the wayside or become generally accepted and move a subject forward. For nearly 30 years Ptitsyn advocated the concept of the 'molten globule' as a key intermediate in protein folding. This was a truly seminal idea — it breathed life into a seemingly intractable field, and as the concept developed it stimulated debate among a whole generation of experimentalists and theoreticians. Opinions still differ as to the semantic value of the terminology, and some members of the folding community use the words with evident anguish. But Ptitsyn's fundamental idea, that proteins can adopt compact structures without the close-packed side-chain interactions characteristic of native proteins, is now implicit in virtually every discussion of protein folding.

Oleg Ptitsyn was born in Leningrad in 1929, and attended the university there as a student of physics. He had a doctorate by the age of 25, and he became a prominent figure in the field of polymer theory through his work at the Institute of High Molecular Weight Compounds of the Academy of Sciences of the USSR in Leningrad. But by the 1960s he had become fascinated by the ubiquitous polymers of the natural world, the proteins. As he later stated, "If a physicist wants to be useful in molecular biology he basically has the choice between establishing the structure of a biological macromolecule or trying to understand the physical basis of a biological process" (*Trends Biochem. Sci.* 20, 376–379; 1995). He had chosen the second way, adding "the main difficulty is finding a process that, while being biologically important, does not involve biochemical reactions, but rather is based purely on physical laws". He concluded that protein folding was the problem waiting to be solved. Nearly 40 years ago this was radical stuff indeed.

To begin research in this new area, Ptitsyn and others founded the Institute of Protein Research in the small scientific settlement of Pushchino, some 70 miles from Moscow. This institute became a world-famous centre for interdisciplinary research. And as a result of the general dispersal of Russian scientists over the past few years, its members have had an especially wide influence.

At Pushchino he allowed his students



and younger colleagues a tremendous amount of intellectual freedom, while maintaining a supportive interest in their activities. It is said that every graduate student he supervised achieved a doctorate, a record of which he was justly proud.

In the early 1970s, Ptitsyn speculated that the protein-folding problem might be made much simpler if a polypeptide chain folds first into a rather flexible state with the native-like mutual positioning of helices and sheets, but without the intricate and detailed packing of the various side-chains found in a fully native protein. There was no experimental evidence for this proposal at the time, and the paper describing the idea passed rapidly through the offices of several journals before it was published (*Dokl. Akad. Nauk. SSSR* 210, 1213–1215; 1973). Shortly afterwards, such evidence began to emerge from studies of protein denaturation in various laboratories. But Ptitsyn was keen to explore his idea himself, and he introduced the experimental strategy of applying a combination of physical methods to search for this new state of proteins. Progress was rapid, and the name 'molten globule' soon emerged (coined in fact by Akiyoshi Wada in Japan). The Ptitsyn laboratory subsequently made the major advance of identifying species in kinetic experiments that fitted his definition of a molten globule, hence relating this state to the mechanism of the folding process itself.

From 1992, Ptitsyn himself spent much of his time as a visiting scientist at the National Institutes of Health in Bethesda, although he also maintained an active group in Pushchino, and indeed collaborations around the globe. To the end of his life he continued to refine his ideas about the molten-globule concept, and to extend it to an ever-widening range of phenomena. He speculated, for example, that such a state could be important in translocation of proteins across membranes, and that its tendency to aggregate could make it a target for the molecular chaperones that help prevent aggregation in the cell. And he was one of the first to link protein misfolding events to disease.

In his later years, however, his greatest enthusiasm was reserved for the topic of molecular evolution. His view was that residues that are important for folding, as well as for function, should be conserved during evolution. By collaborating with other theoreticians and biologists, he obtained the beginnings of the case for this hypothesis. As he said: "Folding represents a bridge between the scientific principles on which the laws of physics are based and the results of the evolutionary pressures under which the character of biology has developed ... For many years we could only discuss this issue in very general terms ... now we are beginning to understand" (*Curr. Opin. Struct. Biol.* 9, 89–91; 1999). He was tasting at last the fruits of the seeds planted all those years ago in the physics laboratory at Leningrad.

Oleg Ptitsyn was a gentle, kindly person, whose diminutive and bustling figure was familiar around the conference and lecture halls of the world. His voice was unmistakable, and following his conversation was a challenge to even the best of linguists. But the effort was never wasted, and his words never contained any of the malice that can diminish the achievements of the greatest of men. He died on 22 March, just before he was due to give a lecture at the University of Warwick, during one of his frequent trips to Britain. It was as he would have wished. He died, as he had lived, earnestly engaged in the practice of science, and looking forward to intense discussions about his latest ideas.

Christopher M. Dobson and R. John Ellis

Christopher M. Dobson is at the Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK.
e-mail: chris.dobson@chem.ox.ac.uk

R. John Ellis is in the Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK.
e-mail: je@dna.bio.warwick.ac.uk

Rocketing from the Galaxy Bazaar

Indian rockets were once the best in the world, and gave the Duke of Wellington a shock he never forgot. But India's social system prevented such technology from fuelling the kind of scientific revolution seen in the West.

Roddam Narasimha

Rockets helped the Mysore army to achieve a famous victory over the British in 1780. The army was led by Hyder Ali, a bold officer who had become the effective ruler of the state, and his son Tipu. The battle is celebrated in a mural at the summer palace in Tipu's capital, Sri-rangapattana. "The fortunes of the English in India had fallen to their lowest water-mark," said the British historian Sir Alfred Lyall, writing in 1914 about this battle in the second Anglo-Mysore War.

A celebrated victim of such a rocket attack was Colonel Arthur Wellesley (later Duke of Wellington and the hero of Waterloo). In the fourth Anglo-Mysore War of 1799, Wellesley suffered a nasty encounter in a mango grove just outside Sri-rangapattana. He lost his way, several of his men were killed and the rest retreated in disorder. This incident had an indelible effect on Wellington, for even late in life he would revert to it with his own "explanations", presumably to counter what his detractors hinted was a blot on his career.

The Rocket Corps in the Mysore army was 5,000 strong in Tipu's time. His rocketmen were skilled in adjusting the elevation of the rocket depending on its size and the distance to the target, and they launched rockets rapidly using a wheeled cart with ramps. Tipu was a 'technology buff', and promoted the manufacture of rockets and other novel devices in areas of his towns often called *Tara-mandalpet* (which translates loosely as Galaxy Bazaar, probably named after the spectacular fireworks known as a star cluster). The rockets' range was typically 2.4 kilometres, an outstanding performance for the time, attributable chiefly to the iron employed for the casing. Indian iron and steel had long been about the best in the world, and permitted increased



Technology buff: Tipu, painted in about 1800.

bursting pressures and hence higher propellant packing density. (European rockets still used some kind of pasteboard.)

The British were so impressed by these rockets that they soon began a vigorous technology programme led by Colonel William Congreve. Several Indian rocket cases were sent to Britain for analysis. In 1801–02, Congreve confirmed with tests that the biggest sky-rockets then available in London had a range less than half that of the Mysore rockets. At the Royal Laboratory at Woolwich Arsenal, he tested various combinations for propellant, and developed a series of rockets with a stout iron case, and iron hoops on one side making it easier to fix the stabilizing stick.

In 1804 he published *A Concise Account of the Origin and Progress of the Rocket System*. Reasoning on the basis of Newton's third law of motion, he recognized that the rocket did not suffer from the recoil that made cannons so difficult to use on ships. In 1806 a rocket

attack on Boulogne, where Napoleon had assembled forces to take war to British soil, set the town on fire, and ended French plans for a cross-Channel expedition. This success was followed by the use of rockets in various other wars in Europe, and in the United States in the War of 1812, when rockets were responsible for the fall of the city of Washington.

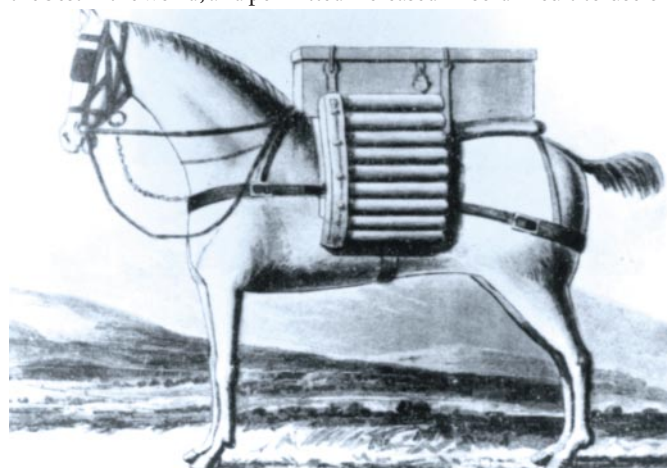
One major reason for interest in this episode is that it occurred during a time of global transition in geopolitics, economics and technology. Clearly, even in the late eighteenth century there were several Indian products technologically superior to Western equivalents, and this was recognized by both sides. But the British effort that followed had the sophistication of research and development today. Scientific principles were applied, designs made, products developed and tested, and all of this was carefully documented — a process alien to Indians of that time. The Indian rockets were well-made but not standardized, being the creation of traditional artisans.

Interestingly, the Indian literature on such technological subjects is rather scarce — even in metallurgy, where the tradition extends over more than three millennia in India. Some of the available texts were written with great authority. A thirteenth-century author said: "I describe for the benefit of the world what I have done myself or observed with my own eyes, not something recorded only from hearsay or from a teacher's instruction." But another exhorted the reader to "guard [this knowledge] with determination, as you would the privacy of your mother": knowledge was handed down only to selected pupils and family members.

Although a certain kind of practical information has always been closely protected around the world, the prevailing social system in India seems to have encouraged great secrecy. The interdependence among the castes, a strong feature of the system that contributed to its remarkable stability, was sustained in part by an 'allocation' of different knowledge subsystems to different castes. So, the craftsmen making the best rockets in the world had no contact with scholars.

Apart from the various causes that have been widely discussed to explain why the scientific and industrial revolutions took place in Europe and not in, say, India, I suggest two others. First, India, being already wealthier, did not need these revolutions. And, second, Indian intellectual property resided in distinct communities that guarded it more jealously and for longer than elsewhere. □

Roddam Narasimha is at the National Institute of Advanced Studies, Bangalore 560 012, India.



Rocket horse: Congreve's six-pound 'service rockets', pictured on a pack horse in 1813, were said to be more effective at frightening the enemy with their noise than at inflicting harm.

Mitochondria and germ-cell death

In birds and mammals, most of the female germ cells are destroyed before fertilization in a process known as atresia, reducing the population of cells to a small fraction of that present in early fetal life. We suggest that this death of germ cells can be interpreted as a developmental solution to the accumulation of mutations (Müller's ratchet) in mitochondria, an idea that is supported by comparative analysis. Atresia in effect therefore removes oocytes carrying mutant mitochondria¹.

In humans, the germ cells are derived from primordial germ cells, which are conspicuous in the developing zygote by the third week after conception. After sexual differentiation, the number of germ cells increases through mitotic division from a few hundred to several million, reaching its peak in humans after five months². The population is then reduced from around five million to one million cells by programmed cell death of the oocytes. Thus atresia, rather than ovulation, is the fate of most follicles³.

Animal mitochondrial genes undergo more mutation than nuclear homologues⁴, and their polymorphism increases significantly with age⁵. If recombination occurs in vertebrate mitochondrial DNA at all, it is likely to be relatively rare. Taken together, these observations make mitochondria particularly vulnerable to the accumulation of mutations across generations.

Figure 1 Phylogenetic regressions for the mitochondrial bottleneck. Relationships between the minimum number of germline mitochondria and the number of offspring (**a**) and the percentage of atretic follicles (**b**). Data were logarithmically transformed to make variances across distantly related groups roughly equal. We used a generalized least-squares model⁶ to determine the degree of association between the number of mitochondria and the number of offspring and level of atresia among developing oocytes. The significance of the regression is evaluated using a log-likelihood ratio (zero covariance over optimal covariance) assuming a chi-squared distribution of likelihood values with one degree of freedom. **a**, Data correspond to: 1, *Meiopriapulus* sp.; 2, *Homo sapiens*; 3, *Macaca mulatta*; 4, *Rattus* sp.; 5, *Sus scrofa domestica*; 6, *Lacerta viridis*; 7, *Drosophila melanogaster*; 8, *Phragmatomosa lapidosa*; 9, *Varroa jacobsoni*; 10, *Elenchus japonicus*; 11, *Rana pipiens*; 12, *Mytilus edulis*; 13, *Raillietiella aegypti*; 14, *Ascaris* sp.; 15, *Arbacia* sp. **b**, 1, *Mytilus edulis*; 2, *Rana pipiens*; 3, *Drosophila melanogaster*; 4, *Raillietiella aegypti*; 5, *Fundulus heteroclitus*; 6, *Lacerta viridis*; 7, *Rattus* sp.; 8, *Cavia* sp.; 9, *Sus scrofa domestica*; 10, *Macaca mulatta*; 11, *Homo sapiens*. Data are from published sources; for several species, several counts of mitochondrial numbers are available. Further details are available from the authors.

The elimination of deleterious mitochondrial genomes before fertilization will be most important in those species producing a small number of offspring over their lifetime. This is because each individual from a small litter accounts for a large fraction of a parent's total reproductive effort⁶, so high-quality mitochondria (and thereby offspring) are required to ensure the survival and viability of descendants. Species that have a very high fecundity can afford to lose some of their offspring, as the death of low-quality individuals will tend to remove deleterious mitochondrial genotypes from the population, leaving behind enough high-quality survivors to reproduce.

The mitochondria responsible for founding the population of organelles within a single diploid individual pass through a severe bottleneck during the maturation of the germ cells⁷. The extent of this bottleneck varies according to species and sex: in the males of most species, it is effectively zero. The bottleneck is achieved by inhibiting mitochondrial replication during cell division. The immediate consequence of the bottleneck is to reduce the genetic variability in a cell and to amplify the differences between cells and individuals, restoring genetic homoplasmy among descendants⁸. Genetic differences between germ cells are necessary for them to compete. A mechanism has recently been suggested by which competition might be mediated directly

through oxidative phosphorylation, namely the relationship between cytochrome *c* and important apoptotic pathways⁹. Furthermore, the bottleneck is roughly coincident with the timing of atresia in mammals, invertebrates and reptiles.

To test whether atresia does serve to reduce the accumulation of mutations, we collected data from several species on the minimum number of mitochondria in developing oocytes, the fraction of atretic follicles in the developing ovaries, and the number of offspring produced, drawing on a range of studies and original sections.

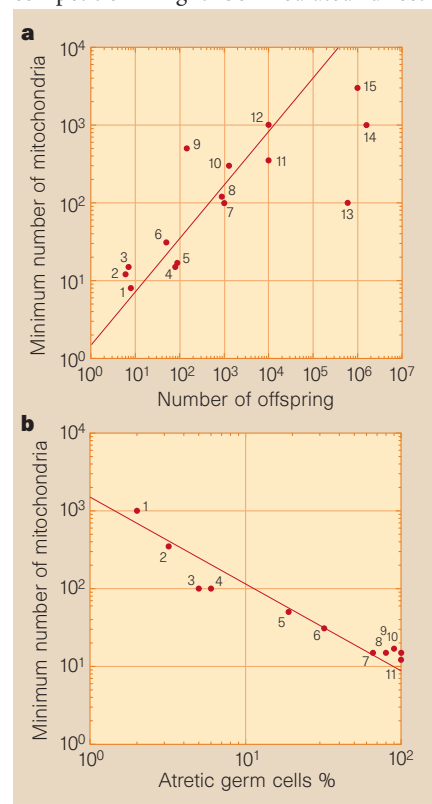
The phylogenetic correlation coefficient¹⁰ for the regression of the number of mitochondria on the number of offspring is shown in Fig. 1a. There is a significant positive correlation, showing that species producing small litters provide the developing oocytes with a smaller number of mitochondria. This amplifies any genetic differences among oocytes, leading to competition and removal of inferior cells and establishes a correlation between the extent of the bottleneck and the number of egg cells removed. The phylogenetic regression coefficient for the regression of the number of mitochondria on the fraction of atretic follicles is shown in Fig. 1b. This negative correlation indicates that the number of mitochondria per cell remains low for species in which a large number of germ cells enter into atresia. This was predicted by our hypothesis, given that high levels of intercellular variance will be required for competition. Our data emphasize that atresia should be viewed as a continuously varying trait across animal groups.

Several other explanations have been put forward for atresia. It may act as a means of regulating hormone production by ovarian follicles¹¹, but this theory cannot explain intraspecific discrepancies in atresia. It may be a local mechanism for eliminating mutations in the nuclear DNA of the egg cells, although oocytes remain at the 4C stage (and hence are clonal) until the moment of fertilization¹². Or finally, it could act to regulate offspring numbers when resources are scarce, although the number of non-atretic follicles remains far greater than the potential number of offspring¹³.

Whichever hypothesis is correct, our comparative analyses demonstrate that the severity of the mitochondrial bottleneck and the total number of offspring produced are strong predictors of the fraction of atretic cells.

David C. Krakauer*, Alex Mira†

*Institute for Advanced Study,
Princeton, New Jersey 08540, USA
e-mail: krakauer@ias.edu



†Department of Zoology, University of Oxford,
South Parks Road, Oxford OX1 3PS, UK

- Jansen, R. P. S. & de Boer, K. *Mol. Cell. Endocrinol.* **145**, 81–88 (1998).
- Baker, T. G. *Proc. R. Soc. Lond. B* **158**, 417–433 (1963).
- Tsafiri, A. & Braw, R. H. *Oxf. Surv. Reprod. Biol.* **6**, 226–265 (1984).
- Lynch, M. *Mol. Biol. Evol.* **13**, 209–220 (1996).
- Corral-Debrinsky, M. et al. *Nature Genet.* **2**, 324–329 (1992).
- Stearns, S. C. in *The Evolution of Sex and its Consequences* (ed. Stearns, S. C.) 337–349 (Birkhauser, Basel, 1987).
- Hauswirth, W. W. & Laipis, P. J. *Proc. Natl Acad. Sci. USA* **79**, 4686–4690 (1982).
- Bergstrom, C. T. & Pritchard, J. *Genetics* **149**, 2135–2146 (1998).
- Shimizu, S., Narita, M. & Tsujimoto, Y. *Nature* **399**, 483–487 (1999).
- Pagel, M. *Zool. Scripta* **26**, 331–348 (1997).
- Curry, S. S. *Biology of Ovarian Follicles in Mammals* 268–273 (Springer, Berlin, 1985).
- Mira, A. J. *Theor. Biol.* **194**, 275–287 (1998).
- Mendez de la Cruz, F. R. et al. *Funct. Ecol.* **7**, 535–540 (1993).

Compensation for wind drift by bumble-bees

In his classic studies on honeybee navigation, von Frisch had to rely on qualitative visual observations of the bees' flight paths, but nevertheless reached the surprising conclusion that bees seem to anticipate lateral wind drift and compensate by flying in shallow curves on the upwind side of their intended tracks^{1,2}. We have investigated wind compensation^{1–3} with much greater precision by using radar^{4,5} to record the flight trajectories of individual bumble-bees (*Bombus terrestris* L.) foraging over arable farmland⁶. Flights typically covered distances of 200 to 700 metres, but bees maintained direct routes between the forage areas and their nests, even in winds with a strong cross-track component. Some bees overcompensated slightly, as described by von Frisch, but most stayed on course by heading partly into the wind and moving obliquely over the ground.

How did the bumble-bees know how far to turn off course to achieve the correct track to their destinations? Several species of Hymenoptera are known to use a Sun compass for navigation^{1,8}, and if, as seems likely, bumble-bees share this ability⁹, we propose that a simple strategy to keep on track in cross-winds would be for them to adjust their headings until the direction of ground image movement over their retinæ (the optical flow¹⁰) occurred at the angle relative to the Sun's azimuth that corresponded to their intended tracks. This mechanism of controlling track direction by direct comparison of two optical directional cues not subject to parallax does not require the bees to assess wind speed or direction, or to compute the combination of heading and air speed required to stay on track. Support for a mechanism based on optical flow comes from visual observations that honeybees flying over water begin to drift with the wind¹¹.

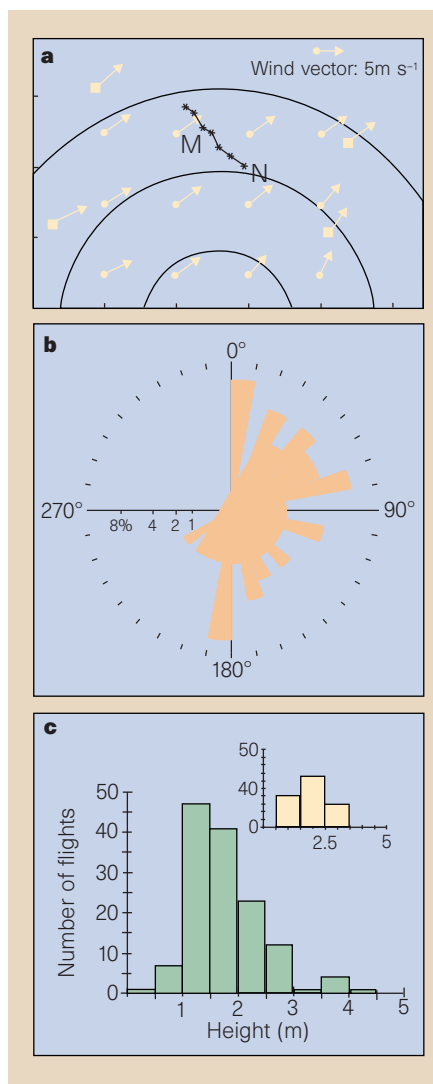


Figure 1a shows an example of cross-wind flight compensation behaviour, which was typical of the 53 return flights examined (Fig. 1b). In still air (where air speed is the same as ground speed), the mean air speed of foragers was $7.1 \pm 0.43 \text{ m s}^{-1}$ ($n=8$), and they flew about 2 m above the ground. In windier conditions, they flew lower on slow upwind tracks and higher in faster downwind movements (Fig. 1c, inset), as though they were adjusting their heights of flight to maintain a preferred rate of optical flow. Results from still air indicate that, if this were the case, the preferred rate for the visual field directly beneath the bees was about $7 \text{ m s}^{-1}/2 \text{ m} = 3.5 \text{ rad s}^{-1}$. We therefore used this value to estimate the height of each flight (Fig. 1c), and subtracted the wind vector at this height from the bees' ground vector to yield air speed and heading.

The overall average air speed found in this way was $7.3 \pm 1.2 \text{ m s}^{-1}$ (s.d.) ($n=74$) in June, and $6.2 \pm 1.2 \text{ m s}^{-1}$ ($n=63$) for July and August, and upwind flight was about 1 m s^{-1} faster than downwind flight in both study periods. The average air speed for June was greater than the fastest air speed (7 m s^{-1}) predicted for bumble-bee

workers, implying a power output exceeding the 180 W kg^{-1} maximum currently attributed to bee flight muscle¹³.
J. R. Riley*, D. R. Reynolds*, A. D. Smith*, A. S. Edwards*, J. L. Osborne†, I. H. Williams†, H. A. McCartney†
*NRI Radar Unit, University of Greenwich, North Site, Leigh Sinton Road, Malvern, Worcestershire WR14 1LL, UK
e-mail: jriley@nri radar.demon.co.uk
†Department of Entomology and Nematology, IACR-Rothamsted, Harpenden, Hertfordshire AL5 2JQ, UK

Figure 1 Estimation of air speed and heading. **a**, A bumble-bee approaching its nest (N) on a cross-wind flight trajectory; asterisks show its position at 3-s intervals. Radar range rings are 115 m apart. North is vertical. Squares are anemometry stations and arrows from them indicate the wind vector 2.7 m above the ground, averaged over the flight; arrows from circles indicate the windfield inferred by interpolation of these vectors. The vertical gradient of the wind speed up to 8 m above ground was measured by four anemometers on a mast (M). From the ground speed we estimate that the bee flew about 2 m above ground level. Subtracting the wind vector at the track centre and at this height (3.8 m s^{-1} towards 54°) from the ground speed vector (6.7 m s^{-1} towards 135°) indicated that the bee's air speed was 7.2 m s^{-1} , and its heading 167° : it was laying off into the wind by 32° from its homeward track. **b**, Circular histogram⁷ of $\theta = |\text{track direction} - \text{wind direction}|$ for bees flying on straight tracks directly to the nest (end-to-end distance/along-track distance > 0.85), $n=53$. Radial scale shows the percentage of tracks falling within each 10° bin, where $\theta=0^\circ$ corresponds to bees flying directly downwind; angles other than 0° and 180° indicate some cross-wind compensation. Wind speeds range from 0.6 to 7.2 m s^{-1} . **c**, Distribution of flight heights, assuming that height = ground speed/3.5 for all flight trajectories ($n=137$). The distribution is similar to that based on our visual assessment that upwind flights occurred between 0.5 and 1.5 m , cross-wind between 1.5 and 2.5 m , and downwind between 2.5 and 3.5 m (inset). This was not surprising, as our tower data showed that mean wind speed varied (as expected¹²) with the logarithm of height, so above 1 m it changed only gradually. The experiments were carried out in June and July/August 1996.

workers, implying a power output exceeding the 180 W kg^{-1} maximum currently attributed to bee flight muscle¹³.

J. R. Riley*, D. R. Reynolds*, A. D. Smith*, A. S. Edwards*, J. L. Osborne†, I. H. Williams†, H. A. McCartney†

*NRI Radar Unit, University of Greenwich, North Site, Leigh Sinton Road, Malvern, Worcestershire WR14 1LL, UK

e-mail: jriley@nri radar.demon.co.uk

†Department of Entomology and Nematology, IACR-Rothamsted, Harpenden, Hertfordshire AL5 2JQ, UK

- von Frisch, K. *Dance Language and Orientation of Bees* (Oxford Univ. Press, London, 1967).
- von Frisch, K. & Lindauer, M. *Naturwissenschaften* **42**, 377–385 (1955).
- Lindauer, M. in *Insect Flight* (ed. Rainey, R. C.) 199–216 (Blackwell, Oxford, 1976).
- Riley, J. R. et al. *Nature* **379**, 29–30 (1996).
- Riley, J. R. et al. *J. Insect Behav.* **11**, 287–296 (1998).
- Osborne, J. L. et al. *J. Appl. Ecol.* (in the press).
- Batschelet, E. *Circular Statistics in Biology* (Academic, London, 1981).
- Wehner, R. *Annu. Rev. Entomol.* **29**, 277–298 (1984).
- Wellington, W. G. *Science* **183**, 550–551 (1974).
- Esch, H. E. & Burns, J. E. *J. Exp. Biol.* **199**, 155–162 (1996).
- Heran, H. & Lindauer, M. *Z. Vergl. Physiol.* **47**, 39–55 (1963).
- McCartney, H. A. & Fitt, B. D. L. in *Advances in Plant Pathology* Vol. 3, *Mathematical Modelling of Crop Disease* (ed. Gilligan, C. A.) 107–143 (Academic, London, 1985).
- Cooper, A. J. Thesis, Univ. Cambridge (1993).

Origin of life from apatite dating?

Mojzsis *et al.*^{1,2} reported the carbon-isotope composition of carbonaceous inclusions in grains of apatite from sediment sequences of Akilia island, southwest Greenland, that are more than 3,850 million years (Myr) old. The $\delta^{13}\text{C}$ values measured by Mojzsis *et al.*¹ led them to conclude that these carbonaceous materials are evidence of early life 3,850 Myr ago. But if other isotopes (U–Pb and Pb–Pb) are used, the apatites are estimated at just $1,504 \pm 336$ (2 σ) and $1,459 \pm 160$ (2 σ) Myr old. This value is consistent with measurements of 1,600–1,700 Myr for Rb–Sr mineral isochrons on biotites³ and 1,670 Myr for K–Ar muscovite⁴ from Amitsoq gneiss in the region. We conclude that, about 1,500 Myr ago, these apatites in Akilia island experienced a metamorphic event of about 600 °C (estimated by the closure temperature of the U–Pb system^{5,6}).

Mojzsis *et al.*¹ measured the carbon-isotope composition of graphite inclusions in grains of apatite in banded-iron formations (BIFs) from Akilia island. Their observed $\delta^{13}\text{C}$ values ranged from –20 per mil (‰) to –50‰, using a PDB standard, indicating a biogenic origin. The BIFs are older than 3,850 Myr (ref. 2), but the age of the apatite housing the graphite material was not determined. Here we present measurements of the U–Pb age of apatites from closely related samples.

We cast the sample chip of the Akilia BIFs (approximately 1 × 1 cm) into epoxy-resin disks with several grains of standard apatite and polished them until they were exposed through their mid-sections. The sample apatites are about 20–30 µm in size and show similar texture to those reported by Mojzsis *et al.*¹. We focused a primary beam of about 2.5 nA O₂[–] to sputter an area of apatites 20 µm in diameter, and extracted the positive secondary ions using 10 kV. We found no isobaric interferences in the mass range over ²⁰⁴Pb and ²⁰⁸Pb at a mass resolution of 5,800. We obtained the ²³⁸U/²⁰⁶Pb ratios from the observed ²³⁸U⁺/²⁰⁶Pb⁺ ratios by calibration using an empirical quadratic relationship between ²⁰⁶Pb⁺/²³⁸U⁺ and ²³⁸U¹⁶O⁺/²³⁸U⁺ ratios of standard. The experimental details of the apatite U–Pb analysis and calibration of data are given elsewhere⁷.

Eleven spots on seven individual apatite grains indicate that U concentrations vary significantly from 22.8 to 132 p.p.m. and do not show any correlation with ²⁰⁶Pb/²⁰⁴Pb and ²⁰⁷Pb/²⁰⁴Pb ratios. A correlation diagram of ²³⁸U/²⁰⁴Pb and ²⁰⁶Pb/²⁰⁴Pb ratios of the Akilia apatites is shown in Fig. 1.

A least-squares fit using the York method gives the ²³⁸U–²⁰⁶Pb* isochron ages of $1,504 \pm 336$ (2 σ ; mean square of weighted deviates (MSWD) = 6.4). A correlation diagram of ²⁰⁴Pb/²⁰⁶Pb and ²⁰⁷Pb/²⁰⁶Pb ratios yields the ²⁰⁶Pb*–²⁰⁷Pb* isochron ages of $1,459 \pm 160$ (2 σ ; MSWD = 1.2). Both ages agree well with each other and are younger than the $3,860 \pm 10$ Myr of ref. 2 for BIFs. This suggests either that the apatites in the BIFs grew about 1,500 Myr ago, or that they grew earlier than that but were subsequently affected by recrystallization, and/or diffusive exchange with the environment, which reset the U–Pb system of the samples. If the apatites were formed during metamorphism, biogenic carbon could have been introduced during the event.

The Akilia association, including the BIFs and Isua supracrustals, are the oldest known components of the Archaean craton of Greenland and were affected by several metamorphic events after their formation 3,800 Myr ago (ref. 2). The latest event recorded in rock samples is the injection of basic dykes and crustally derived granitic sheets about 1,600 Myr ago, possibly coupled with the anatectic reheating⁸. Biotite from all types of gneisses in the area gives a Rb–Sr isochron age of 1,600–1,700 Myr ago (ref. 3), which is within the error of the isochron ages of the Akilia apatites.

If these apatites were formed 3,800 Myr ago, graphite inclusions within grains of apatites have also experienced a thermal event of around 600 °C about 1,500 Myr ago. The isotope composition of the carbonaceous inclusions might have been altered by the event. It has been suggested⁹ that the oxidation of carbonaceous matter during metamorphism could alter the initial ¹³C value from –10‰ to –35‰ if temperatures are higher than 500 °C and if an oxidizing system is provided. In iron formation, an oxidizing system seems a reasonable suggestion.

The young U–Pb ages on the apatites, indicating closure for Pb in an event as late

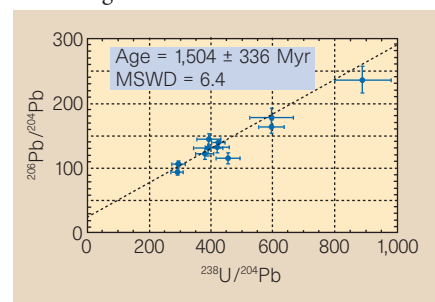


Figure 1 Correlation diagrams of ²³⁸U/²⁰⁴Pb and ²⁰⁶Pb/²⁰⁴Pb ratios of Akilia apatites. Errors are shown at the 1 σ level. The dotted line shows the best fit by the York method. In the calculation of this ²³⁸U/²⁰⁴Pb–²⁰⁶Pb/²⁰⁴Pb diagram, an error correlation of $r = 0.8554$ was used, derived from the correlation coefficient between $\delta(^{238}\text{U}/^{204}\text{Pb})/(^{238}\text{U}/^{204}\text{Pb})$ and $\delta(^{206}\text{Pb}/^{204}\text{Pb})/(^{206}\text{Pb}/^{204}\text{Pb})$.

as 1,500 Myr, and the theoretical possibility that the light signature is a product of fractionation since 3,850 Myr, might in isolation throw doubt on the proposal¹ that the Akilia BIF hosted life when deposited at around 3,850 Myr. However, as stated by Mojzsis *et al.*¹, the petrographic association of the apatite and carbon-rich material is well known from unmetamorphosed BIF with microfossils, and also from modern biological observations. This was considered an equally important line of evidence for biological activity when the Akilia BIF was deposited.

In conclusion, when we seek evidence of the earliest life on Earth, we need to find an apatite with a U–Pb closure age apparently older than 3,500 Myr; with a pattern of rare-earth elements that indicates a biogenic signature; and with carbonaceous inclusions that contain isotopically light carbon.

Yuji Sano, Kentaro Terada, Yoshio Takahashi, Allen P. Nutman

Department of Earth and Planetary Sciences, Hiroshima University, Kagamiyama 1-3, Higashi-Hiroshima 739, Japan
e-mail: ysano@ipc.hiroshima-u.ac.jp

1. Mojzsis, S. J. *et al.* *Nature* **384**, 55–59 (1996).
2. Nutman, A. P., McGregor, V. R., Friend, C. R. L., Bennett, V. C. & Kinny, P. D. *Precamb. Res.* **78**, 1–39 (1996).
3. Baadsgaard, H., Lambert, R. St J. & Krupicka, J. *Geochim. Cosmochim. Acta* **40**, 513–527 (1976).
4. Pankhurst, R. J., Moorbath, S., Rex, D. C. & Turner, G. *Earth Planet. Sci. Lett.* **20**, 157–170 (1973).
5. Cherniak, D. J., Lanford, W. A. & Ryserson, F. J. *Geochim. Cosmochim. Acta* **55**, 1663–1673 (1991).
6. Krogstad, E. J. & Walker, R. J. *Geochim. Cosmochim. Acta* **58**, 3845–3853 (1994).
7. Sano, Y., Oyama, T., Terada, K. & Hidaka, H. *Chem. Geol.* **153**, 249–258 (1999).
8. Kalsbeek, F., Bridgwater, D. & Boak, J. *Pap. Grönlands Geol. Unders.* **100**, 73–75 (1980).
9. Eiler, J. M., Mojzsis, S. J. & Arrhenius, G. *Nature* **386**, 665 (1997).

Mojzsis *et al.* reply — Sano *et al.* report a ²⁰⁷Pb/²⁰⁶Pb date of $1,459 \pm 160$ Myr for apatite grains 25 µm in diameter, containing inclusions of isotopically light carbon, from a granulite grade BIF (typically 70% quartz, 10% magnetite, 20% mafic phases and sulphides) from Akilia island, Greenland, that was lithified before 3,850 Myr (refs 1,2). To Sano *et al.*, this suggests either that the apatites “grew about 1,500 Myr ago, or that they grew earlier than that but were subsequently affected by recrystallization, and/or diffusive exchange”.

Two lines of evidence support the latter idea. The first is that radiogenic Pb is not quantitatively retained in apatite around 25 µm in diameter³ at temperatures above 400 °C, during regional slow cooling (1–10 °C per Myr), or during a transient heating event that persists for several million years. It has long been known^{4,5} that this region experienced a thermal excursion to temperatures exceeding 400 °C about 1,500 Myr ago. Thus, the Pb–Pb apatite age reported

by Sano *et al.* of approximately 1,500 Myr is the expected result.

Second, this rock, which is essentially a quartzite, has no intrinsic capacity to form apatite by metamorphic reactions following diagenesis. Thus, the apatite present in this rock must be original (formed before 3,850 Myr) unless phosphate-bearing fluids were subsequently introduced metasomatically. Although no apatite has been recognized in the BIF that does not contain graphite inclusions, no apatites with graphite inclusions have been found in the immediately adjacent, encompassing gneisses^{1,6,7}.

Sano *et al.* are concerned that if "the apatites were formed during metamorphism, biogenic carbon could have been introduced". Again, given that the encompassing rocks do not contain graphitic inclusions with isotopically light carbon¹, this could not have occurred after 3,850 Myr.

"If these apatites were formed 3,800 Myr ago," Sano *et al.* allow, "graphite inclusions within grains of apatite have also experienced a thermal event of around 600 °C" (actually 400 °C; see <http://oro.ess.ucla.edu/sjm-nature.html>) "about 1,500 Myr ago". They posit that the "isotope composition of the carbonaceous inclusions might have been altered by the event". However, the extremely slow self-diffusion rates of carbon in graphite⁷ (less than 10^{-39} cm² s⁻¹ at 600 °C) precludes significant diffusive carbon-isotope exchange.

Could the isotopically light carbon ($\delta^{13}\text{C}$; 35‰) in the apatite inclusions¹ be abiogenic? Sano *et al.* seem to prefer a Rayleigh distillation mechanism that was previously⁸ raised to illustrate the unlikelihood of such a process occurring. However, such a process has not been documented in the geological record, and Sano *et al.* do not specify how abiogenic carbon of ($\delta^{13}\text{C}$ exceeding 10‰; ref. 9) could be fractionated by about 30‰ to mimic a biogenic signature. Under metamorphic conditions, the relevant mineral assemblage (quartz, magnetite, pyroxene, amphibole, pyrrhotite, apatite and graphite) coexists with H₂O-bearing fluid containing both CH₄ and CO₂ (ref. 10). The previous calculations⁸ did not consider the participation of CH₄, which mitigates against this type of isotope fractionation because it preferentially partitions ¹²C, leaving a residue that becomes isotopically heavier with time, not lighter¹¹. Thus, the role for Rayleigh fractionation proposed by Sano *et al.* is both improbable and unsupported by their Pb isotope data for apatite. Finally, Sano *et al.* imply that the rare-earth-element signatures of apatites can distinguish between a biogenic and abiogenic origin, but there is no basis for such a suggestion in this context.

In conclusion, the data reported by Sano *et al.* are not pertinent to the formation age of the host apatite and are not relevant to

the mode of origin or preservation of the isotopically light carbonaceous inclusions they contain. In our view, the occurrence of this geochemical association in rocks of sedimentary origin is best explained as the product of biological activity^{1,12} more than 3,850 Myr ago^{1,2}.

S. J. Mojzsis*†, T. M. Harrison*, G. Arrhenius‡, K. D. McKeegan*, M. Grove*

*Department of Earth and Space Sciences, University of California at Los Angeles, Los Angeles, California 90095-1567, USA
e-mail: sjm@argon.ess.ucla.edu

†Institute for Pure and Applied Physical Sciences, University of California at San Diego, La Jolla, California 92093-0360, USA

‡Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0220, USA

1. Mojzsis, S. J. *et al.* *Nature* **384**, 55–59 (1996).
2. Nutman, A. P., Mojzsis, S. J. & Friend, C. R. L. *Geochim. Cosmochim. Acta* **61**, 2475–2484 (1997).
3. Cherniak, D. J., Lanford, W. A. & Ryerson, F. J. *Geochim. Cosmochim. Acta* **55**, 1663–1673 (1991).
4. Baadsgaard, H., Lambert, R. S. & Krupicka, J. *Geochim. Cosmochim. Acta* **40**, 513–527 (1976).
5. Pankhurst, R. J., Moorbath, S., Rex, D. C. & Turner, G. *Earth Planet. Sci. Lett.* **20**, 157–170 (1973).
6. Thrower, P. A. & Meyer, R. M. *Phys. Status Solidi* **47**, 11–37 (1978).
7. McGregor, V. R., & Mason, B. *Am. Mineral.* **62**, 887–904 (1977).
8. Eiler, J., Mojzsis, S. J. & Arrhenius, G. *Nature* **386**, 665 (1997).
9. Deines, P. & Gold, D. P. *Geochim. Cosmochim. Acta* **37**, 1709–1733 (1973).
10. Ohmoto, H., & Kerrick, D. M. *Am. J. Sci.* **277**, 1013–1044 (1977).
11. Mojzsis, S. J. & Arrhenius, G. *J. Geophys. Res.* **103**, 28495–28511 (1998).
12. Schidlowski, M. *Nature* **333**, 313–318 (1988).

Nicotine withdrawal and road accidents

Waters *et al.*¹ suggested that the effects of nicotine withdrawal increase the chances of an accident at work. The second Wednesday of March is 'No Smoking Day' (NSD) in the United Kingdom. If the 'accident effect' is real, increased numbers of other types of accident would be expected on NSD. I have investigated this possibility by using data from the UK national STATS19 accident reporting system², which records details of personal injury accidents occurring on public roads in which at least one

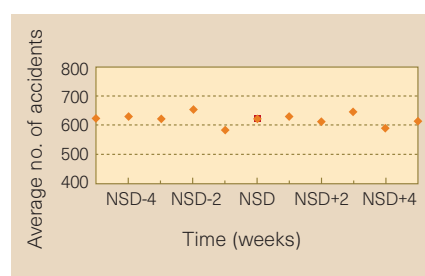


Figure 1 Average number of accidents in 1988–97 on Wednesdays around 'No Smoking Day' (NSD).

road vehicle, or a vehicle in collision with a pedestrian, is involved, and which becomes known to the police within 30 days of its occurrence. Damage-only accidents are not included. The date of the accident is also recorded.

I compared the average number of road accidents occurring over the past 10 years (1988–97) on NSD with the average for the Wednesday before and the Wednesday after NSD. Table 1 shows the means and standard deviations of these measures for NSD week and the weeks before and after it. Paired-sample Student's *t*-tests (one-tailed) did not indicate that there were significantly more accidents on NSD than on the previous Wednesday ($t = -1.36$, $P = 0.10$) or the following one ($t = -0.39$, $P = 0.35$).

A visual inspection of the data for the five Wednesdays before and after NSD (Fig. 1) shows no obvious indication that the effects of nicotine withdrawal on NSD led to an increased number of road accidents when compared with the other Wednesdays in February and March. The only feature to stand out is the decrease in accidents between the last Wednesday in February and the first Wednesday in March in 7 of the 10 years examined. No explanation for this effect has been established.

Similar analyses of the severity of accidents and the age and sex of the drivers involved also provide no evidence of an increased number of accidents on NSD.

Jackie Knowles

Safety and Environment Resource Centre, Transport Research Laboratory, Old Wokingham Road, Crowthorne, Berkshire RG45 6AU, UK

1. Waters, A. J., Jarvis, M. J. & Sutton, S. R. *Nature* **394**, 137 (1998).
2. Department of the Environment, Transport and the Regions *Road Accidents in Great Britain 1997* (HMSO, London, 1998).

Table 1 Average number of road accidents in 1988 to 1997

Day	One week before NSD		NSD week		One week after NSD	
	Average no.	s.d.	Average no.	s.d.	Average no.	s.d.
Monday	593.4	63.6	598.8	84.5	627.0	99.0
Tuesday	579.8	122.3	605.6	85.9	606.4	93.3
Wednesday	583.2	98.2	620.7	58.4	628.1	47.8
Thursday	655.3	106.5	605.0	61.5	639.0	85.9
Friday	712.1	70.3	719.0	67.3	781.3	90.0
Saturday	624.3	71.8	649.0	61.4	633.1	103.6
Sunday	453.3	61.7	502.9	60.8	488.3	77.8

s.d., Standard deviation.

Deviating from the line

A Beautiful Mind

by Sylvia Nasar

Faber & Faber/Simon & Schuster: 1998.
459 pp. £17.99, \$25 (hbk); £9.99, \$16 (pbk)

Karl Sigmund

In 1994, John Forbes Nash Jr received the Nobel Prize in Economics for a one-page note published in 1950. It introduced what today is called the Nash equilibrium, arguably the most basic concept, not just for game theory, but for every model of a social situation where the consequences, for each agent involved, depend not only on their own actions but also on those of the others. In a Nash equilibrium, none of the 'players' has an incentive to change strategy as long as the other players keep to theirs.

Game theory, with its stated ambition to address economic interactions of interdependent decision-makers, had been founded by John von Neumann and Oskar Morgenstern. In 1948, the second edition of their monumental book *Theory of Games and Economic Behavior* (Princeton University Press) dominated and in fact circumscribed the field: utility theory, zero-sum games, coalitions ... The 20-year-old Nash, whose education in economics consisted of one undergraduate course, had some ideas which, as he mentioned in his Nobel autobiography, "deviated somewhat from the 'line' [as in party political line] of von Neumann and Morgenstern's book".

Von Neumann was bluntly dismissive of the young man: "This is trivial, you know. That's just a fixed-point theorem." And, in a way, von Neumann was right. On the other hand, his own celebrated maximin theorem reduces to a simple corollary of Nash's existence proof, and is valid for two-person zero-sum games only; these include plenty of parlour games, but cover painfully few instances of real economic interactions. Nash's "rival approach" has incontestably won the day.

Classics in Game Theory, edited by Harold Kuhn (Princeton University Press, 1996), begins with three papers written by Nash before he was 25. With masterful poise and lucidity, they redefined the field by introducing Nash equilibria, the Nash solution of the bargaining problem and the 'Nash program' relating cooperative and non-cooperative game theory (for example, basing coalitions on negotiations).

In the following 50 years, the subject expanded enormously, but with hindsight it can be seen that half-sentences in his unpublished PhD thesis and footnotes "attributed to Dr Nash" had anticipated recent developments in some of the most active areas, like experimental economics and evolutionary game theory. For instance, the classical inter-



A journey of sadness and grandeur: Nash's prime years were devastated by schizophrenia.

pretation of a Nash equilibrium relies on the rationality of the players, who ought to be able to deduce a solution consistent with their diverse interests (she thinks that I think that she thinks ...); but Nash also proposed a "mass action" interpretation, based "on a large population of players using empirical information to arrive at a stable frequency distribution of strategies". Many years later, this approach, which dispenses with rationality, was rediscovered by theoretical biologists and led to the concepts of unbeatable strategy (William D. Hamilton) and evolutionarily stable strategy (John Maynard Smith) which dominate current thinking on adaptation by natural selection and percolate through the social sciences.

Fundamental as Nash's economic insights were, they used only 'easy' mathematics — what von Neumann had called trivial. This may have been the reason Nash turned away from game theory and attacked some of the technically hardest mathematical problems. He showed, in particular, that every Riemannian manifold (a curved space which looks, to any near-sighted inhabitant, like Euclidean space) could be smoothly embedded, without altering any distances, in a much higher-dimensional Euclidean space.

Nash also obtained basic results on the

existence, uniqueness and continuity of solutions to nonlinear partial differential equations — equations of supreme importance to applied mathematics (fluid dynamics, for instance, and electromagnetic fields). In each case, he overcame staggering difficulties by relentlessly pushing ahead along completely unexpected lines.

This work might well have earned him a Fields medal (the mathematical equivalent of the Nobel, but restricted to people under 40 years old). By the time Nash was 28, it seemed only a matter of time: his reputation for brilliance was solidly established. *Fortune* magazine featured him as "one of the brightest new stars in New Maths". Here was a most unusual young man, tall and aloof, with strikingly good looks, married to a smart, beautiful woman.

Schizophrenia usually strikes at a younger age. In Nash's case, earlier symptoms may have been attributed to the eccentricities to be expected of a genius, the passing effects of fierce concentration and mental strain, or the arrogant self-absorption of an immature wunderkind. But slowly it became clear that Nash was sinking into a state of paranoid delusion. He gave up his faculty position at the Massachusetts Institute of Technology, and attempted to become a political refugee in Europe. "I started to think that I was a man of great religious importance ... I began to hear something like telephone calls in my head, from people opposed to my ideas ... The delirium was like a dream from which I never seemed to awake."

With the exception of a brief remission in the mid-1960s, during which Nash resumed first-class work on fluid dynamics and singularities, "an interlude of, as it were, enforced rationality", his prime years were devastated by the terrible illness. After several involuntary hospitalizations, Nash eventually drifted back to Princeton, a tragic, silent shell of his former self, haunting the campus as the "Phantom of Fine Hall", searching for secret meanings in numbers and leaving bizarre messages on blackboards.

But, "gradually, I began to intellectually reject some of the delusionally influenced lines of thinking which had been characteristic of my orientation". In the 1980s, Nash, who had resumed his life with his ex-wife Alicia, managed to recover from an illness that had been deemed incurable, and to return to scientific work. In his autobiographical note, Nash defiantly grounds his "hopes of [still] being able to achieve something of value" on having gone through "a sort of vacation", namely "the gap period of 25 years of partially deluded thinking". The enthusiastic support from a new generation of game theorists — Jörgen Weibull and

Ariel Rubinstein, in particular — helped to pave the way for the Nobel prize, a belated tribute to one of the most influential intellectual figures of our time.

This story, a modern-day version of the biblical legend of Job, is bound to attract biographers. Sylvia Nasar, an economics correspondent from the *New York Times*, was offered a visiting position at Princeton's Institute for Advanced Study to prepare her book. Her excursions into mathematics are kept to a minimum, and make little pretence at explaining Nash's discoveries. She attempts to describe the Klein bottle (a one-sided surface) without offering a figure, and does not bother to draw the board of Hex, a clever game invented by Nash as a graduate student. Her book does not provide many details on the illness either. In addition, Nash chose not to cooperate with his biographer — the arrests, deportations and enforced commitments must have been too agonizing to retrace.

Despite these disadvantages, Nasar has succeeded in writing an admirable book that grips from beginning to end. Based on hundreds of interviews, her depiction of the small tribe of top mathematicians is uncannily true to life, a masterpiece of oral history; and her inside reports from the offices of the RAND corporation ("thinking the unthinkable") or the conclaves of the Swedish Academy of Sciences are gems of investigative journalism. Most importantly, her sober, honest and humane portrayal of Nash is an absorbing experience, and fully conveys the sadness and grandeur of his fate. □

Karl Sigmund is at the Institute of Mathematics, University of Vienna, Strudlhofgasse 4A-1090, Vienna, and the International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, Laxenburg, Austria.

● *A Beautiful Mind* was short-listed for the Rhône-Poulenc Prize for Science Books (see *Nature* 399, 654; 1999).

When ideas become tollbooths

Owning the Future

by Seth Shulman

Houghton Mifflin: 1999. 240 pp. \$25

A. P. Simonds

There was a time when the phrases 'marketplace of ideas' and 'free flow of information' were considered closely connected, if not synonymous. But the trends surveyed by science journalist Seth Shulman suggest that, as these phrases turn from metaphor to literal description, their mutual antagonism becomes unavoidably apparent. At the end of the day, the wedding of business and knowledge may be captured by the blunt aphorism:

there's no such thing as a free lunch.

The new landscape of intellectual property has attracted wide attention, and not just from law firms, research labs and application-development shops. A contentious literature flows out of every corner of academia, from business schools to English departments. Yet, ironically, the subject has scarcely risen to the point of visibility in the national policy debates of most industrial democracies.

Beneficiaries of strong patent protection frequently present themselves as free-market individualists. Shulman describes an eye surgeon who had acquired a patent on a procedure discovered by accident and essentially similar to techniques already used by others. The surgeon defended his claim to payment of thousands of dollars in annual royalties from every other surgeon performing the procedure by noting that, after all, "medicine is a capitalist endeavour".

Yet patent protection derives not from a private right to property but from the power of the state to grant monopoly privileges on the grounds that the public's interest in "the Progress of Science and useful Arts" (to use the language of the US Constitution) overrides the individual's right to practise his trade unfettered by government constraint. Defining the nature, extent and validity of that public interest is surely essential to the business of a democratic polity.

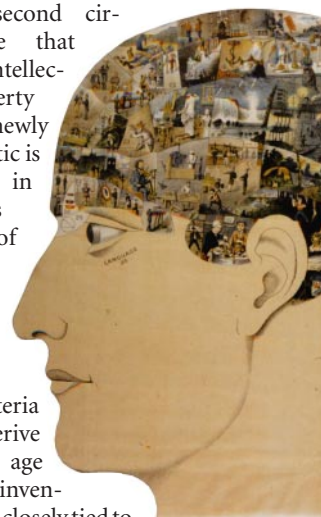
Shulman seeks to stimulate this debate by describing a series of striking and, at least on their legal surface, egregiously expansive claims to patent protection for software routines, hybridized or recombinant organisms, and natural genetic resources. Two features of the late twentieth-century environment (rather than a simple rise in greed and incompetence) appear chiefly responsible for the observed trends. The first is that, in a knowledge-based economy, incentives to aggressively pursue and defend the broadest possible patent protection are more powerful than ever before. This is partly because personal and national wealth depends, to a far greater extent than previously, on the control of knowledge resources. But it is also the result of a classic collective-action dilemma: each company's ability to defend itself against patent claims brought by others depends, in part, on the portfolio of patents it can bring to the table in support of cross-licensing agreements. The cost of litigating these arrangements may undermine the productivity of all parties, but none of them can do anything except push forward with vigorous and expansive new claims. Shulman quotes Jonas Salk's 1954 answer to Edward R. Muro's question about who would control the new polio vaccine: "There is no patent. Could you patent the Sun?" A half-century later, this answer would be much more difficult to sustain.

The second circumstance that makes intellectual-property claims newly problematic is evident in Shulman's selection of examples. Patent

offices apply standards and criteria that derive from an age in which inventions were closely tied to discrete, material things. This does not mean that patenting a process is some novelty of the information age. (The very first patent issued in the United States was for a process used to produce potash.) But it does mean that concepts crucial to the distinction between what can and cannot, should and should not be afforded protection as intellectual property ("invention", "product of nature", "non-obvious", "language", "procedure", etc.) cannot be applied in a straightforward and unproblematic way to phenomena such as software codes, business procedures or DNA sequences.

Lawyers are quick to point out that nothing in the law authorizes ownership of an idea; all that can be protected is its "tangible expression" (copyright) or its "practical application" (patent). What Shulman shows, however, is that loopholes and ambiguities in the standards for assigning protection can and are being exploited to turn ideas into "tollbooths" that allow private appropriation of the tools of thought and which obstruct or impede the very innovation and public dispersal of knowledge that provide the public-policy justification for intellectual property in the first place.

This happens in several interrelated ways. Overly broad patents can close off competing research avenues. (Shulman offers the example of a 1994 European patent granting Agracetus exclusive rights to "any and all genetically altered soybeans", whether created by the method specified on their application or any other.) Abstract, essentially conceptual anticipations of devices allow a skilful ideas merchant to reap enormous licensing rewards once actual products are brought to market by others. (Shulman mentions Jerome Lemelson, awarded more US patents than anyone other than Thomas Edison, as the unchallenged master of this art.) Patents that push at the boundaries of "obvious" and "algorithm" (both non-patentable) leave programmers at constant risk of inadvertent infringement. Accidents of priority and



initiative allow one patent applicant to capture control over innovations that issue from the work of large numbers of others and, not infrequently, significant expenditures of public funds.

Shulman's account leaves unanswered questions; indeed, some of his examples are still in the process of review and litigation. His arguments are far from conclusive, and this is not the place to look for academic or legal rigour. But the tocsin he sounds is loud, clear and compelling. The new forms of private appropriation of knowledge must be brought squarely into the centre of public debate, lest we find ourselves forced to resurrect and rephrase A. J. Liebling's famous remark about freedom of the press in a far more expansive form: "freedom of thought is guaranteed only to those who own some." □

A. P. Simonds is in the Department of Political Science, University of Massachusetts at Boston, Boston, Massachusetts 02129, USA.

Dos and don'ts, whys and wherefores

Ever Since Adam and Eve: The Evolution of Human Sexuality

by Malcolm Potts and Roger Short
Cambridge University Press: 1999. 368 pp.
\$74.95, £47.50

Ruth Mace

Books about evolution and sex are beginning to outnumber even those titles on how to give birth smiling or how to teach your baby to read before it is three. They have been written by everyone from journalists to entomologists, and I have never got further than the first few paragraphs of most of them. So the prospect of another tome with a naked Eve seducing a naked Adam on the front cover did not fill me with great anticipation. As the mother of two sons, I find myself wishing that someone would shed more light on what Cain did to Abel rather than what Adam did to Eve. But if you have not read the others (or if you have, but still want more), then just read this one.

What distinguishes this book is that the authors actually have some relevant experience in the area of human reproduction. They bring to the book not only a professional awareness of reproductive health and fertility around the world, but also insights from evolution, anthropology, psychology, the history of medicine and, for that matter, art. It is not just about sex. It covers everything from sex in primates to marriage in humans, child-rearing, ageing, sexually transmitted disease and the

population explosion. It is peppered with arresting anecdotes, and illustrations which range from works by the great masters to pictures out of medical texts (where they probably should have remained).

A recurring theme is how we are adapted to a hunter-gatherer lifestyle. Much of what followed the origin of agriculture, and then urbanization, is viewed as plunging us out of our evolutionary depth. The authors are not very keen on the unnatural. They don't like bottle-feeding (on the basis of some evidence) or day care for toddlers (on the basis of no evidence at all). Thankfully, natural childbirth is not given any hallowed status and natural fertility is viewed as unwelcome.

But if 'natural' means what our forager ancestors did, then we have very little idea of what that was. They may have been peaceful and monogamous, with carefully spaced children, as the !Kung of southern Africa appear to have been. Or they may have been polygynous, infanticidal and highly fertile, as were the Ache of Paraguay. What is natural to a !Kung may be abhorrent to an Ache, not to mention disastrous for his reproductive success. Spacing births is good if you want to minimize infant mortality, but if evolution minimized infant mortality nobody would have more than one child. Evolution maximizes reproductive success, and if you want to maximize reproductive success you may have to suffer some infant mortality along the way.

Some of the most important insights into child health gained from sociobiology relate to the fact that it is sometimes in the reproductive interest of parents to harm their own children — and, as Malcolm Potts and Roger Short show, they do. Sex biases in parental investment, based on differences in the reproductive returns from investing in each sex, provide an example of this. A picture of a woman from a

patrilineal Asian society, breast-feeding a healthy boy and bottle-feeding his emaciated twin sister, is presented as an example of 'breast is best'. Such a clear illustration of infanticidal neglect is shocking. On another page, a little boy pictured dying of starvation while his mother breast-feeds two healthy twins is described as a victim of a short birth interval. This he may be, but birth intervals also reflect parental choice. This boy is from a predominantly matrilineal area of southern Africa, where the sex-biases in reproductive returns on parental investment are the reverse of those seen in patrilineal societies, and where levels of malnutrition and mortality in boys clearly exceed those seen in girls. The maximization of reproductive success did not stop in the Pleistocene.

Potts and Short provide some alarming statistics on population growth and make a convincing case for the importance of contraception, backed up by abortion, as fundamental to achieving small families. The implication is that everybody wants contraception, and that if everybody had it, everybody would have small families.

This is the model on which family planning initiatives have operated for decades, and it may be true for much of the world. But the demographic transition to low fertility started in eighteenth-century Europe when no contraceptives were available, and has yet to start in other areas of Africa and the Middle East, where contraceptives are on offer everywhere. Social scientists insist that contraceptive use would be increased if women had more power, neglecting the fact that many women are happy with their big families.

Demographers are giving up their hunt for a single, socio-economic correlate of the onset of fertility decline. But evolutionary ecologists have much to say about variation in parental investment, and thus also family



Different: an advertisement for contraception.

size. The omission of Darwinian interpretations of contemporary reproductive patterns from this book reflects an omission in the whole field of human demography. I hope it is beginning to be redressed as human evolutionary ecology provides a new perspective on these important issues.

Evolutionary approaches are valuable, not because they tell us what is natural or what is good, but because they help us understand why people do what they do. So, more evolution please. □

Ruth Mace is in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK.

Sensation and explanation

Going Inside: A Tour Round a Single Moment of Consciousness

by John McCrone

Faber & Faber: 1999. 368 pp. £20

John C. Marshall

There are two profound mysteries about human consciousness. First, why are so many books and articles being published on the topic? Second, why does most of the information in these works concern what we can know and do without being conscious of it? There is a simple answer to these questions: we know nothing of scientific value about consciousness (except in the very important sense in which the anaesthetist uses the term). If ever there was a perfect domain of investigation for scholars working on the social construction of science, 'consciousness' must be it: the way in which the notion has bobbed in and out of (more or less) respectable psychological inquiry deserves serious study.

Having said all this, I can now admit that John McCrone's *Going Inside* is far superior to the vast majority of recent tomes on cognitive neuroscience for the general reader. McCrone rounds up the usual suspects, but at least he does so with some care. He begins with Stephen Kosslyn's work on visual imagery, a good choice because it incorporates an informative mix of traditional psychological experimentation and functional brain imaging. And he ends with consciousness, a less good choice, as the chapter basically concludes with the admission that we have no idea "how a lump of flesh like the brain could light up with the inner glow of subjective experience". Apparently, a young Australian philosopher, David Chalmers, became rich and famous by making this somewhat negative (albeit correct) claim at a meeting in Arizona.

Luckily, in the middle sections McCrone just gets on with the job of describing the recent history and current status of some



To be three or ...

This self-portrait of René Magritte appears in *Consciousness* by J. Allan Hobson (Scientific American Library, \$34.95, £23.95). The book explores the relationship between the mind and the brain from the premise that "we are our brains and our brains are us". Thus, it describes brain structures and functions that are now understood to be fundamental to conscious experience. Here is what the book

has to say about the Magritte self-portrait.

"He is, of course, Magritte, but only in the sense that Magritte is Everyman, the essential person with all the unessential details left out. The consciousness of the bowler-hatted man is generic, having the same formal features every day and — by the light of the moon — every night, in a reliably state-dependent manner."

interesting issues in the brain sciences. There are particularly fine descriptions of work on attention, expectation and prediction — from the psychophysics of ball games (baseball and cricket) to the positive shift in the electrical activity of the brain (P 300) that signals surprise.

My only quibble is that McCrone totally misses out on the war work of the UK Medical Research Council's Applied Psychology Unit in Cambridge. The original impetus to study the mind/brain as a predictive system surely came from investigations of gunnery by Kenneth Craik and his colleagues. Likewise, Claude Shannon's mathematical theory of communication led Cambridge psychologists to realize that the 'stimulus' was not merely the event that actually occurred, but rather the selection of that event out of the set of stimuli that might have occurred. Brain scientists, including those interviewed for this book, have a regrettable tendency to ignore what was discovered by techniques other than their own.

By contrast, the story that leads from Donald Hebb's synaptic learning rule to 'neural nets' and the whole connectionist movement is dealt with at length. But equal stress is placed on modern functional brain-imaging techniques, and McCrone lucidly

describes how the physiological measures they provide can help us to understand the neurobiology of object-recognition, memory and the emotions. Where he is less clear about what is actually going on (the organization of language in the brain, for example), his perplexity probably reflects confusion in the research he is describing.

Two themes recur with great regularity: 'dynamic system' and 'brain plasticity' are clearly the catch-phrases of the decade. The data to which these phrases direct us are now well established. Yes, the output of the brain is constantly adapting to a huge range of endogenous and exogenous variables. And yes, even the adult brain can (to some extent) rearrange its topography after injury. But we still have far to go in understanding how any cognitive function is actually represented in the brain. McCrone concludes his discussion of brain imaging with the very reasonable remark that, "having splashed out on the machines and hyped up the expectations, the neuroscience community will feel it has to deliver something". □

John C. Marshall is in the Department of Clinical Neurology, University of Oxford, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK.

Determining computational complexity from characteristic 'phase transitions'

Rémi Monasson*, Riccardo Zecchina†, Scott Kirkpatrick‡, Bart Selman§ & Lidror Troyansky||

* CNRS-Laboratoire de Physique Théorique, 24, Rue Lhomond, 75231 Paris Cedex, France

† The Abdus Salam International Centre for Theoretical Physics, Strada Costieri 11, 34100 Trieste, Italy

‡ IBM, Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA

§ Computer Science Department, Cornell University, Ithaca, New York 14853, USA

|| Institute of Computer Science, The Hebrew University, Jerusalem 91904, Israel

Non-deterministic polynomial time (commonly termed 'NP-complete') problems are relevant to many computational tasks of practical interest—such as the 'travelling salesman problem'—but are difficult to solve: the computing time grows exponentially with problem size in the worst case. It has recently been shown that these problems exhibit 'phase boundaries', across which dramatic changes occur in the computational difficulty and solution character—the problems become easier to solve away from the boundary. Here we report an analytic solution and experimental investigation of the phase transition in K -satisfiability, an archetypal NP-complete problem. Depending on the input parameters, the computing time may grow exponentially or polynomially with problem size; in the former case, we observe a discontinuous transition, whereas in the latter case a continuous (second-order) transition is found. The nature of these transitions may explain the differing computational costs, and suggests directions for improving the efficiency of search algorithms. Similar types of transition should occur in other combinatorial problems and in glassy or granular materials, thereby strengthening the link between computational models and properties of physical systems.

Many computational tasks of practical interest are surprisingly difficult to solve even using the fastest available machines. Such problems, found for example in planning, scheduling, machine learning, hardware design, and computational biology, generally belong to the class of NP-complete problems^{1–3}. NP stands for 'non-deterministic polynomial time', which denotes an abstract computational model with a rather technical definition. Intuitively speaking, this class of computational tasks consists of problems for which a potential solution can be checked efficiently for correctness, yet finding such a solution appears to require exponential time in the worst case. A good analogy can be drawn from mathematics: proving open conjectures in mathematics is extremely difficult, but verifying any given proof (or solution) is generally relatively straightforward.

The class of NP-complete problems lies at the foundations of the theory of computational complexity in modern computer science. Literally thousands of computational problems have been shown to be NP-complete. The completeness property of NP-complete problems means that if an efficient algorithm for solving just one of these problems could be found, one would immediately have an efficient algorithm for all NP-complete problems. However, a fundamental conjecture of modern complexity theory is that no such efficient algorithm exists.

Although NP-complete problems are believed to require exponential time to solve in the worst case, the typical-case behaviour is difficult to characterize. Yet, such typical-case properties are most relevant in practical applications. Fu and Anderson⁴ first conjectured a deep connection between NP-complete problems and models studied in statistical physics⁵. More recently, it was discovered (refs 6–9) that NP-complete problems can exhibit phase transition phenomena, analogous to those in physical systems, with the hardest problems occurring at the phase boundary. (For a collection of papers and a review, see refs 10, 11.) The exact relationship between phase transition phenomena and computational properties has remained unclear. For example, phase

transitions are also observed in computationally 'easy' problems (that is, ones that are not NP-complete).

Here we establish the nature of the relationship between phase transition phenomena and typical-case computational complexity. More specifically, we show that when the underlying computational task exhibits a continuous ('second order') phase transition, resource requirements grow only polynomially with problem size, while discontinuous ('random first order'¹²) phase transitions correspond to an exponential growth in resource requirements, characteristic of truly hard computational problems. Our results show that techniques from statistical physics can provide important new insights into computational phenomena, possibly leading to improved solution methods. In addition, computational problems, viewed as many-particle systems, provide models which challenge the assumptions of physics, and may shed light on the behaviour of some highly nonlinear materials now coming under study.

K -SAT: a standard test bed

We focus on the K -satisfiability (K -SAT) problem^{1–3,11}, which is characteristic of the class of NP-complete problems, and is commonly used as a test bed for heuristic algorithms. An instance of the K -SAT problem is defined by a set of N boolean variables, each of which can be assigned 'true' or 'false', and a set of M clauses. Each clause is a logical OR of K variables, where each variable can be negated. The goal is to find an assignment to the variables such that all clauses are satisfied. For example, the 2-SAT formula, $(x \text{ OR } y) \text{ AND } ((\text{NOT } x) \text{ OR } (\text{NOT } y))$, can be satisfied by setting x to true and y to false. In the worst case, for $K \geq 3$, an exponential search through the space of 2^N truth assignments is required in order to find a satisfying assignment or to determine that no such assignment exists, in which case the formula is called 'unsatisfiable'. The K -SAT problem was the first problem shown to be NP-complete¹. A remarkable consequence of the theory of NP-completeness is that a large collection of problems (in fact, several thousand) from many different fields can be represented as instances of the K -SAT

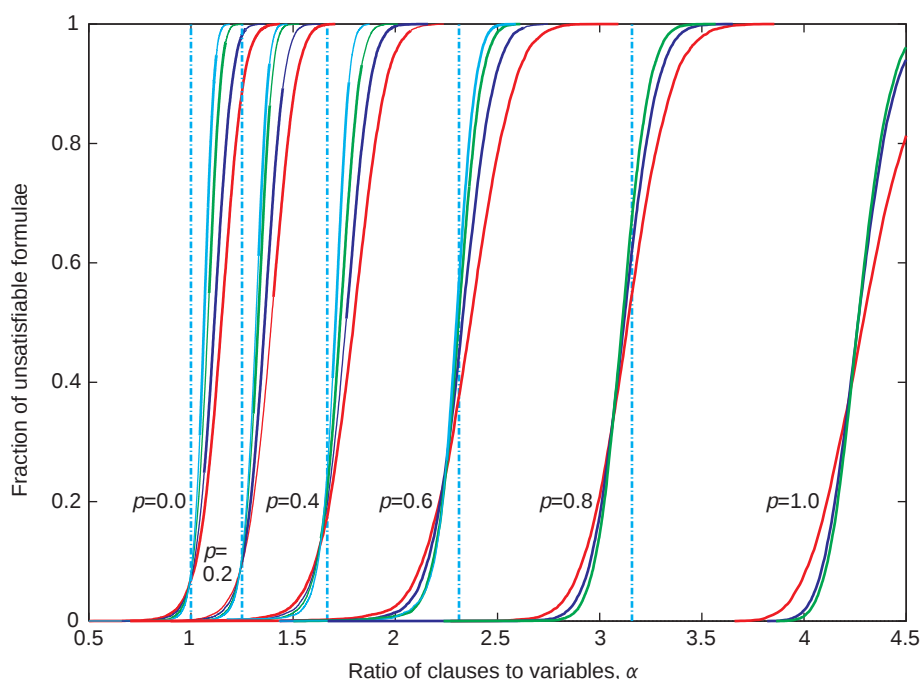


Figure 1 The SAT/UNSAT phase transition for the $(2+p)$ SAT problems for a range of values of p and N . Each curve was generated by considering 10,000–40,000 randomly generated formulae, for a range of values of α . For $p = 0$, N has values of 1,000 (red), 2,000 (blue), 5,000 (green), 10,000 (light blue); for $p = 0.2$, N has values of 1,000 (red), 2,000 (blue), 5,000 (green), 7,500 (light blue); for $p = 0.4$, N has values of 500 (red), 1,000 (blue), 2,000 (green), 5,000 (light blue); for $p = 0.6$, N has values of 250 (red), 500 (blue), 1,000 (green), 1,500 (light blue); for $p = 0.8$, N has

values of 200 (red), 350 (blue), 500 (green); for $p = 1.0$, N has values of 100 (red), 200 (blue), 250 (green). The largest value of N which could be achieved for each value of p was limited by the increasing cost of the calculations as p is increased. We note that $p = 0.0$ corresponds to 2-SAT, and $p = 1.0$ to 3-SAT. The vertical lines mark the thresholds as predicted by the replica symmetric (RS) theory. The transition curves sharpen with increasing N .

problem². In a sense, the problem lies at the core of computational complexity theory, and the analysis of K -SAT therefore provides insights into a broad range of challenging computational problems.

We study the typical case complexity of K -SAT by considering an ensemble of randomly generated K -SAT formulae. For each formula, we generate M clauses, where each clause is generated by randomly selecting K variables from the set of N variables and negating each selected variable with probability 0.5 to construct the clause. Formulae constructed at random, keeping the ratio of clauses to variables, $\alpha = M/N$, constant as $M, N \rightarrow \infty$, provide a natural ensemble of test problems.

The value of K is important: the 1-SAT and 2-SAT problem can be solved efficiently, that is, in time polynomial in size of the formula (in fact, a linear time algorithm is known¹³). We say that 1-SAT and 2-SAT belong to the class 'P' (polynomial time solvable problems). The 2-SAT problem exhibits phase transition phenomenon at a critical ratio of clauses to variables equal to one ($\alpha_c(2) = 1$). Below this critical ratio almost all formulae are satisfiable (SAT), whereas above this ratio almost all are unsatisfiable (UNSAT)^{14,15}. The intuitive explanation for this phenomenon is that at low values of α , we have relatively few clauses and many variables, which means that the formula is relatively easy to satisfy (the problem is under-constrained). At high values of α , the problem is over-constrained and generally no satisfying assignment exists. The most interesting formulae can be found at the transition from the under-constrained to the over-constrained area. We will make these observations more precise below.

For $K \geq 3$, K -SAT is NP-complete. Using computer experiments, we have shown⁸ thresholds for $K = 3, 4, 5$ and 6. A threshold exists for any value of K (ref. 16), but determining the exact location of the threshold remains a difficult open problem. One reason for the interest in the threshold is the growing recognition that 'easy–hard–easy' patterns of computational cost are characteristic of heuristic search algorithms developed for solving NP-complete problems,

and that the hardest instances occur near phase boundaries^{6–11}. (The easy–hard–easy pattern refers to the facts that (1) at low values of α , it is relatively easy to find a satisfying assignment, (2) at values of α near the threshold, it is most difficult to find a satisfying assignment or show unsatisfiability of the formula, and (3) at high values of α , it becomes again relatively easy to solve the formulae (in this case, by showing them unsatisfiable).) Above the critical ratio, α_c , proving UNSAT is still hard, with a lower bound on the search cost that is exponential in N with a coefficient of N that decreases as a power law in α (ref. 17). Because of the extreme hardness of formulae at the threshold, such formulae provide a useful benchmark for evaluating new search methods.

Interpolating between P and NP

To understand the onset of exponential complexity that occurs when going from a problem in P (2-SAT) to a problem that is NP-complete (3-SAT), we introduce here formulae containing mixtures of 2- and 3-clauses. Consider a random formula with M clauses, of which $(1-p)M$ contain two variables and pM contain three variables, with $0 \leq p \leq 1$. This ' $(2+p)$ -SAT' model smoothly interpolates between 2-SAT ($p = 0$) and 3-SAT ($p = 1$). Figure 1 shows that the threshold from SAT to UNSAT shifts as a function of p from that for 2-SAT to that for 3-SAT. The problem is NP-complete for any value of $p > 0$, as any such instance contains a sub-formula of 3-clauses. Thus, the worst-case complexity is probably exponential, but the typical-case complexity need not be, as we find below.

There is a strong analogy between randomly generated K -SAT problems and disordered materials, alloys or glasses, studied by constructing models whose energy function captures the unsatisfied constraints. In studies of strongly disordered models with conflicting interactions, similar to the randomly negated variables in K -SAT, the replica method is used⁵ to confirm that ordering of an unusual type is possible in the presence of microscopic randomness. We have previously applied replica methods to determine the

characteristics of the 3-SAT problem^{18–20}. Here we report results on the $(2+p)$ -SAT problem, which provides new insights into the typical-case complexity, in particular with respect to the change in going from a computationally tractable (polynomial) problem to an intractable (NP-complete) problem. We hope that these results are useful both to computer scientists, in constructing better algorithms, and to statistical physicists, in their study of disordered systems.

The $(2+p)$ -SAT model can be mapped onto a diluted spin glass model with N spin variables S_i : $S_i = 1$ if the boolean variable x_i is true, $S_i = -1$ if x_i is false. We can associate with each configuration an energy E , or cost-function, equal to the number of clauses violated by that configuration. (Note that a configuration corresponds to an assignment of truth values to the boolean variables.) The replica formalism provides an order parameter describing the statistics of optimal assignments, that is, those ‘ground-state’ assignments that minimize the number of violated (unsatisfied) clauses. Consider an instance of the $(2+p)$ -SAT problem. We use the N_{GS} ground-state configurations to define $m_i = 1/N_{GS} \sum_{g=1}^{N_{GS}} S_i^g$, that is, the average value of spin S_i over all optimal configurations. Clearly, m_i ranges from -1 to $+1$, and $m_i = -1$ means that the corresponding boolean variable x_i is always false in all ground states; conversely, $m_i = +1$ means that x_i is always true in all ground states. The distribution $P(m)$ of all m_i characterizes the microscopic structure of the ground states. The accumulation of magnetizations m around ± 1 represents a ‘backbone’ of almost completely constrained variables, whose logical values cannot vary from solution to solution, while the centre of the distribution $P(m \approx 0)$ describes weakly constrained variables.

The backbone fraction is the order parameter

For $\alpha < \alpha_c$, the solution exhibits a simple symmetry property, usually referred to as replica symmetry, which leads to an order parameter which is precisely the magnetization distribution $P(m)$ defined above. An essential qualitative difference between 2-SAT and 3-SAT is the way that the order parameter $P(m)$ changes at the

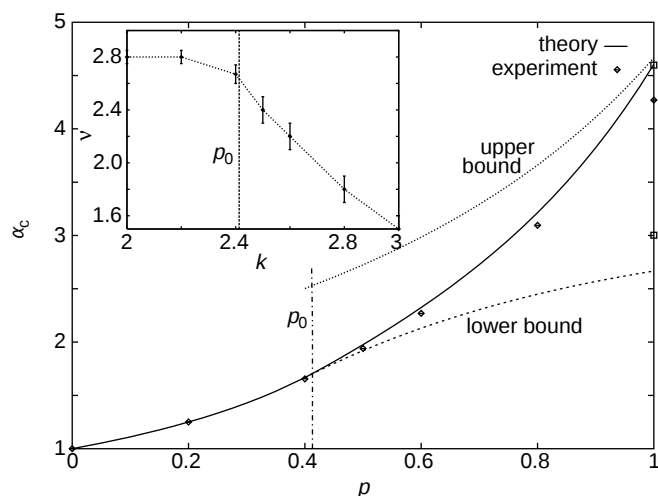


Figure 2 Theoretical and experimental results for the SAT/UNSAT transitions in the $(2+p)$ -SAT model. The vertical line at p_0 (~ 0.41) separates the continuous from the discontinuous transition. The full line is the replica-symmetric theory's predicted transition, believed exact for $p < p_0$. The diamond data points are results of computer experiment and finite-size scaling. The other two lines show upper and lower bounds obtained in ref. 21, while the stronger upper bound due to ref. 32 (see also refs 33, 34), and the best known lower bound, due to ref. 35, are indicated by the two square data points (only known for $p = 1.0$, that is, $K = 3$). Inset, the values of ν (the finite-size scaling exponent) obtained from finite-size scaling analysis. (K is the average number of variables per clause). We note that the finite-size scaling exponent ν now depends on p , dropping roughly linearly to 1.5 at $p = 1.0$.

threshold. This discrepancy can be seen in the fraction $f(K, \alpha)$ of boolean variables which become fully constrained, at and above the threshold. Below the threshold, $f(K, \alpha) = 0$. (Consider adding one clause to a satisfiable formula found below α_c . If there is a finite fraction of backbone spins, there will be a finite probability that the added clause creates an unsatisfiable formula, but we know that the probability of an unsatisfiable formula tends to zero for large N , for $\alpha < \alpha_c$.) For 2-SAT, $f(2, \alpha)$ becomes strictly positive above $\alpha_c (= 1)$ and is continuous at the transition: $f(2, 1^-) = f(2, 1^+) = 0$. By contrast, $f(3, \alpha)$ displays a discontinuous change at the threshold: $f(3, \alpha_c^-) = 0$ and $f(3, \alpha_c^+) = f_c(3) > 0$.

For the mixed $(2+p)$ -SAT model, the crucial issue is therefore to understand how a discontinuous 3-SAT-like transition mechanism may appear when increasing p from zero up to one, and to see how it affects the computational cost of finding solutions near the threshold. Using replica methods, we find a continuous (second-order) SAT/UNSAT transition at $\alpha_c(2+p) = 1/(1-p)$ for $p < p_0$, where p_0 , estimated by various methods, lies between 0.4 and 0.416. This transition point can be calculated by simply ignoring the 3-SAT clauses. (Note that we have $(1-p)M$ 2-SAT clauses, and for 2-SAT $(M/N)_c = \alpha_c = 1$.) It is perhaps somewhat counterintuitive that the 3-SAT clauses do not contribute at all to the location of the phase transition for $p < p_0$ (in a sense, the 2-SAT constraints completely dominate the formulae), but this is consistent with a recent analytical result²¹ derived for our model up to $p = 0.4$. The replica symmetric theory appears to be correct for all $\alpha < \alpha_c(2+p)$ when $p < p_0$, and predicts the threshold correctly, as in the $K = 2$ case.

We have employed finite-size scaling analysis to locate the transition region⁸, in which the likelihood of a formula being UNSAT grows from zero to one, in a region centred around $\alpha_c(2+p)$ and of width proportional to $N^{-1/\nu}$. The values of α_c obtained (Fig. 2) agree with the replica symmetric theory for $p < p_0$. Below p_0 , the exponent ν is roughly constant and close to 2.8, the value found for 2-SAT (Fig. 2, inset). Corrections to scaling could make ν compatible with the value 3 obtained both from mean-field theory and from the recent exact bounds on the growth of the 2-SAT backbone (B. Bollobas, C. Borgs, J. Chayes, J. Kim, D. Wilson, personal communication). This suggests that the $(2+p)$ -SAT transition is in the same universality class as 2-SAT for $p < p_0$. So, for $p < p_0$, the $(2+p)$ -SAT model shares the physical features of random 2-SAT with a continuous phase transition. A plausible hypothesis is that $\nu = 3$ for all $p < p_0$.

For $p > p_0$, the transition becomes discontinuous and the replica symmetric transition can only provide an upper bound for the true $\alpha_c(2+p)$. The replica symmetric (RS) theory correctly predicts a discontinuous appearance of a finite fraction of fully constrained variables which jumps from 0 to f_c when crossing the threshold $\alpha_c(2+p)$. However, values of both $f_c(2+p)$ and α_c are overestimated; for example, for $p = 1$, $\alpha_c^{RS}(3) \approx 4.60$ and $f_c^{RS}(3) \approx 0.94$. Subtracting the terms in $P(m)$ which appear to be splitting away from $m = \pm 1$ reduces that to an estimated $f_c(3) \approx 0.6$ in an approximation beyond replica symmetry, while experiments give $\alpha_c(3) \approx 4.27$ and $f_c(3) \approx 0.4$. A replica symmetry breaking theory will be necessary to predict these quantities. For $p > p_0$, the random $(2+p)$ -SAT problem shares the physical features of random 3-SAT.

Scaling of computational cost

Previous work showed that the cost of running the best heuristics, depth-first search algorithms with backtracking²² increases exponentially with N for any value of $\alpha \geq \alpha_c$ for $K = 3$ (refs 7, 23). The cost reaches its maximum value at $\alpha_{0.5}(K, N)$, the value of α for which the probability of the formula being satisfiable equals 0.5, and we have the largest uncertainty about whether the formula is satisfiable or not, leading to the greatest difficulty for the search algorithm. Our data show that the computational cost at $\alpha_{0.5}(K, N)$

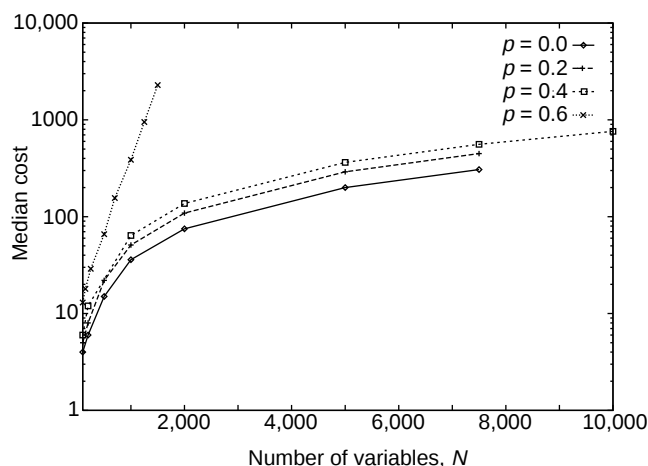


Figure 3 The median computation cost (number of backtracks) of proving a formula SAT or UNSAT in the $(2 + p)$ -SAT model for a range of values of p . We used a state-of-the-art implementation of the Davis–Putnam search method³⁶ for $p = 0.0, 0.2, 0.4, 0.6$, fixing $\alpha = \alpha_{0.5}$ (the hardest region), over the range of N that could be efficiently searched. The semi-log plot shows an exponential dependence of cost on N for $p = 0.6$ ($p > p_0$), but it remains linear in N for $p \leq p_0$.

scales linearly with N for $p = 0$ through 0.4 , and exponentially (Fig. 3) for $p = 0.6$. These results are again consistent with the observation that $(2 + p)$ -SAT behaves essentially as 2-SAT for $p \leq p_0$, and as 3-SAT for $p > p_0$.

The presence of a ‘backbone’ of frozen spins provides a good explanation for the apparently high cost of heuristic search near the phase boundary. A backtrack search method which proceeds by asserting possible spin values can make early mistakes by setting backbone spins (variables) to incorrect values, and then take a long time to backtrack to correct these mistakes. (The search may have to explore an exponential size subtree with no solutions²⁴.) The positive value of $f(2, \alpha)$ above α_c is consistent with the fact that finding the ‘best’ assignment, that is, one that satisfies the most clauses, is NP-complete for 2-SAT.

Our experiments confirm that the appearance of the backbone is discontinuous (Fig. 4). Above the threshold, the fraction of frozen spins which was found in small samples by exhaustive enumeration of all ground states is relatively insensitive to N . Below the threshold, the fraction of frozen spins decreases rapidly with increasing N . Although the samples that could be studied are small, the results suggest that $f(2 + p, \alpha)$ vanishes below α_c and are consistent with $f(2, \alpha_c^+)$ vanishing while $f(3, \alpha_c^+)$ reaches a limit of slightly more than 0.4 for large N .

Discussion

The random first-order phase transition in K -SAT is unusual because it has a mixed nature: of first order in the order parameter, yet continuous in thermodynamic quantities such as the entropy^{12,25}. This means that while an extensive number of degrees of freedom freeze at the transition, forming the backbone, the remaining ones can support critical fluctuations. The separation of the backbone is a realization of the ‘two components’ (corresponding to slow and fast degrees of freedom) idea sometimes put forward on heuristic grounds to describe the physics of glasses²⁶ or granular materials²⁷. The exponent ν shown in Fig. 2 inset is quite different from the value, 1 , expected for typical first-order transitions, or for the mean-field theories that usually describe models with no spatial dimensionality.

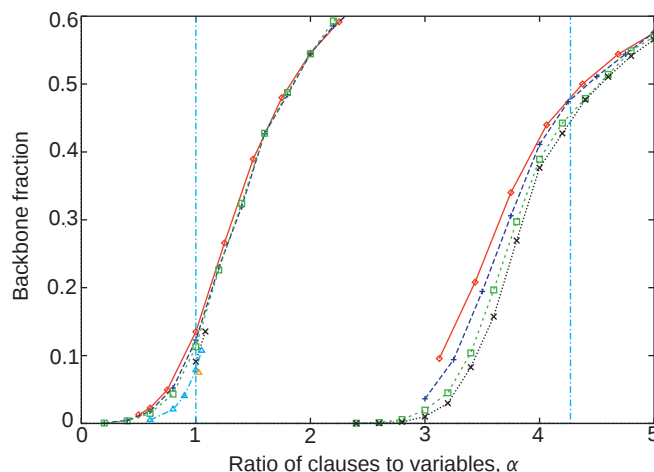


Figure 4 Backbone fractions as a function of α for 2-SAT and 3-SAT. The data were extracted from 3,000–15,000 samples for the 2-SAT cases (on the left) with N values of 20 (red), 30 (blue), 45 (green), 100 (black), 200 (light blue), and 500 (orange). The results for $N \leq 45$ were obtained by exhaustive enumeration, examining all assignments, while those for $N \geq 100$ used a modified Davis–Putnam search procedure. The 3-SAT cases were studied by exhaustive enumeration in 7,500–30,000 samples for N values of 16 (red), 20 (blue), 24 (green), and 28 (black). The vertical lines mark the observed SAT/UNSAT thresholds in the limit $N \rightarrow \infty$. For 2-SAT, data obtained from larger sizes show that the backbone fraction at the threshold tends towards zero.

The clear experimental evidence for the coexistence of a first-order transition with critical fluctuations is, to our knowledge, new with K -SAT, although there have been hints of such a possibility occurring in ‘rigidity percolation’ models^{28,29} intended to capture the mechanical properties of sand and similar granular materials.

We note that conventional first-order transitions, such as the freezing of a solid from its liquid form, do not imply computational complexity. Describing the ground state of a crystal, once the positions of a few atoms are established, requires only the application of symmetry to calculate positions for the remaining atoms with an effort that would be linear in the number of atoms. By contrast, a spin glass may have an infinite number of ground states (in the limit $N \rightarrow \infty$), and no translational symmetry.

The existence of a backbone has also been observed in other combinatorial problems, such as the travelling salesman problem³⁰. It is possible to improve heuristic optimization methods, at least for this problem, by early identification and elimination of the backbone. The remaining problem consists of many separate subproblems, each of which is much less complex to solve³¹. This may prove to be a generally valid approach. Efficient means of finding the backbone would be specific to each problem type, but should nonetheless provide a step forwards in algorithm efficiency. Moreover, many NP-complete problems occurring in practice contain a mix of tractable and intractable constraints. Our results suggest that the run time of search algorithms that exploit as much of the tractable structure as possible may in fact scale polynomially in the typical case. We hope that our work will stimulate other studies to develop a further understanding of the structure of phase transitions taking place in computational problems, as studied in computer science, and in diluted random systems, as studied in statistical physics. Both fields seem to share several basic open problems. □

Received 21 December 1998; accepted 11 May 1999.

1. Cook, S. A. in *Proc. 3rd Annu. ACM Symp. on Theory of Computing* 151–158 (Assoc. Comput. Machinery, New York, 1971).
2. Garey, M. & Johnson, D. S. *Computers and Intractability: A Guide to the Theory of NP-completeness* (Freeman, San Francisco, 1979).
3. Papadimitriou, C. *Computational Complexity* (Addison-Wesley, Reading, 1994).

4. Fu, Y. T. & Anderson, P. W. Application of statistical mechanics to NP-complete problems in combinatorial optimization. *J. Phys. A* **19**, 1605–1620 (1985).
5. Mézard, M., Parisi, G. & Virasoro, M. A. *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
6. Cheeseman, P., Kanefsky, B. & Taylor, W. M. in *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI-91)* (eds Mylopoulos, J. & Reiter, R.) 331–337 (Morgan Kaufmann, San Mateo, California, 1991).
7. Mitchell, D., Selman, B. & Levesque, H. in *Proc. 10th Natl. Conf. on Artificial Intelligence (AAAI-92)* 440–446 (AAAI Press/MIT Press, Cambridge, Massachusetts, 1992).
8. Kirkpatrick, S. & Selman, B. Critical behavior in the satisfiability of random Boolean expressions. *Science* **264**, 1297–1301 (1994).
9. Selman, B., Mitchell, D. & Levesque, H. Generating hard satisfiability problems. *Artif. Intell.* **81**, 17–29 (1996).
10. Hogg, T., Huberman, B. A. & Williams, C. (eds) *Frontiers in problem solving: phase transitions and complexity*. *Artif. Intell.* **81**(I & II) (1996).
11. Hayes, B. Can't get no satisfaction. *Am. Sci.* **85**, 108–112 (1996).
12. Gross, D. & Mezard, M. The simplest spin glass. *Nucl. Phys. B* **240**, 431–452 (1984).
13. Aspvall, B., Plass, M. F. & Tarjan, R. E. A linear-time algorithm for testing the truth of certain quantified Boolean formulas. *Inf. Process. Lett.* **8**, 121–123 (1979).
14. Goerdts, A. A threshold for unsatisfiability. *J. Comput. System Sci.* **53**, 469–486 (1996).
15. Chvátal, V. & Reed, B. Mick gets some (the odds are on his side). *Proc. 33rd IEEE Symp. on Foundations of Computer Science* 620–627 (IEEE Computer Soc. Press, New York, 1992).
16. Friedgut, E. Sharp thresholds of graph properties and the K-SAT problem. *J. Am. Math. Soc.* (in the press).
17. Beame, P., Karp, R., Pitassi, T. & Saks, M. in *Proc. ACM Symp. on Theory of Computing (STOC98)* 561–571 (Assoc. Comput. Machinery, New York, 1998).
18. Monasson, R. & Zecchina, R. Entropy of the K-satisfiability problem. *Phys. Rev. Lett.* **76**, 3881–3885 (1996).
19. Monasson, R. & Zecchina, R. The statistical mechanics of the random K-satisfiability model. *Phys. Rev. E* **56**, 1357–1370 (1997).
20. Monasson, R. & Zecchina, R. Tricritical points in random combinatorics: the 2+p SAT case. *J. Phys. A* **31**, 9209–9217 (1998).
21. Achlioptas, D., Kirousis, L., Kranakis, E. & Krizanc, D. in *Proc. Workshop on Randomized Algorithms in Sequential, Parallel and Distributed Computing (RALCOM 97)*, 1–10 (1998).
22. Davis, M. & Putnam, H. A computing procedure for quantification theory. *J. Assoc. Comput. Machinery* **7**, 201–215 (1960).
23. Selman, B. & Kirkpatrick, S. Critical behavior in the cost of K-satisfiability. *Artif. Intell.* **81**, 273–295 (1996).
24. Gomes, C. P., Selman, B. & Crato, N. in *Principles and Practice of Constraint Programming (CP97)* (ed. Smolka, G.) 121–135 (Lect. Notes in Comput. Sci., Vol. 1330, Springer, Heidelberg, 1997).
25. Kirkpatrick, T. R., Thirumalai, D. & Wolynes, P. G. Scaling concepts for the dynamics of viscous liquids near an ideal glassy state. *Phys. Rev. A* **40**, 1045–1054 (1989).
26. Angell, A. C. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
27. Bouchaud, J.-P., Cates, M. E., Ravi Prakash, J. & Edwards, S. F. Hysteresis and metastability in a continuum sandpile model. *Phys. Rev. Lett.* **74**, 1982–1985 (1995).
28. Jacobs, D. J. & Thorpe, M. F. Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.* **75**, 4051–4054 (1995).
29. Moukarzel, C. & Duxbury, P. M. Stressed backbone and elasticity of random central-force systems. *Phys. Rev. Lett.* **75**, 4055–4058 (1995).
30. Kirkpatrick, S. & Toulouse, G. Configuration space analysis of traveling salesman problems. *J. Phys. (France)* **46**, 1277–1292 (1985).
31. Schneider, J., Froehammer, C., Morgenstern, I., Husslein, T. & Singer, J. M. Searching for backbones—an efficient parallel algorithm for the traveling salesman problem. *Comput. Phys. Commun.* **96**, 173–188 (1996).
32. Kirousis, L., Kranakis, E. & Krizanc, D. in *Proc. 4th European Symp. on Algorithms* 27–38 (Springer, Berlin, 1996).
33. Dubois, O. & Boufkhad, Y. A general upper bound for the satisfiability threshold of random K-SAT formulas. *J. Algorithms* **24**, 395–420 (1997).
34. Kamath, A., Motwani, R., Palem, K. & Spirakis, P. Tail bounds for occupancy and the satisfiability threshold conjecture. *Random Struct. Algorithms* **7**, 59–89 (1995).
35. Frieze, A. & Suen, S. Analysis of two simple heuristics on a random instance of K-SAT. *J. Algorithms* **20**, 312–335 (1996).
36. Crawford, J. & Auton, L. in *Proc. 11th Natl. Conference on Artificial Intelligence (AAAI-93)* 21–27 (AAAI Press/MIT Press, Cambridge, Massachusetts, 1993).

Acknowledgements. S.K. thanks the Laboratories of Theoretical and Statistical Physics at Ecole Normale Supérieure for a visiting professorship in 1998, during which this work was completed. B.S. is an Alfred P. Sloan research fellow and is supported by an NSF Faculty Early Career Development Award. During the early phase of this research, B.S. was at AT&T Bell Laboratories.

Correspondence and requests for materials should be addressed to S.K. (e-mail: kirk@watson.ibm.com).

A confirmed location in the Galactic halo for the high-velocity cloud 'chain A'

Hugo van Woerden*, Ulrich J. Schwarz*, Reynier F. Peletier*†, Bart P. Wakker‡ & Peter M. W. Kalberla§

* Kapteyn Institute, Postbus 800, 9700 AV Groningen, The Netherlands

† Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

‡ Department of Astronomy, University of Wisconsin, Madison, Wisconsin 53706, USA

§ Radio-astronomisches Institut, Universität Bonn, 53121 Bonn, Germany

The high-velocity clouds of atomic hydrogen, discovered about 35 years ago^{1,2}, have velocities inconsistent with simple Galactic rotation models that generally fit the stars and gas in the Milky Way disk. Their origins and role in Galactic evolution remain poorly understood³, largely for lack of information on their distances. The high-velocity clouds might result from gas blown from the Milky Way disk into the halo by supernovae^{4,5}, in which case they would enrich the Galaxy with heavy elements as they fall back onto the disk. Alternatively, they may consist of metal-poor gas—remnants of the era of galaxy formation^{2,6–8}, accreted by the Galaxy and reducing its metal abundance. Or they might be truly extragalactic objects in the Local Group of galaxies^{7–9}. Here we report a firm distance bracket for a large high-velocity cloud, chain A, which places it in the Milky Way halo (2.5 to 7 kiloparsecs above the Galactic plane), rather than at an extragalactic distance, and constrains its gas mass to between 10^5 and 2×10^6 solar masses.

Distance estimates of high-velocity clouds (HVCs) have long been based on models or indirect arguments^{2,10}. The only direct method uses the presence or absence of interstellar absorption lines at the HVC's velocity in spectra of stars at different distances. The presence of absorption shows the HVC to lie in front of the star; absence places it beyond, provided that the expected absorption is well above the detection limit¹¹. Blue stars are best for this process, since their spectra contain few confusing stellar lines. The method requires that metal ions are present in HVCs. Indeed, since suitable spectrographs have become available, Ca II and other metal-ion absorption lines have been found for many HVCs³ in the spectra of background quasars or Seyfert galaxies. Using H I column densities measured at high angular resolution, the metal-ion/H I ratios thus derived provide estimates of expected absorption-line strengths towards stars probing the same HVC.

The HVCs MII–MIII, for which an upper limit to the distance, $d < 4$ kpc, is known^{12,13}, but no lower limit¹³, were thus found to be Galactic, and may even lie in the Galactic disk, as does the tiny object^{14,15} HVC100 – 7 + 100, with $d < 1.2$ kpc, distance from the plane $|z| < 0.14$ kpc. Lower distance limits are known for complex C^{16,17} ($d > 2.5$ kpc, and probably¹⁸ $d > 5$ kpc), cloud 211 = HVC267+ 20 + 215¹⁶ ($d > 6$ kpc), parts of the Anticenter complexes¹⁹ ($d > 0.6$ kpc), and complex H²⁰ ($d > 5$ kpc). Chain A is the first HVC for which both a significant upper and a non-zero lower limit to its distance are known, constraining its location relative to the Galaxy's major components.

HVC-complex A, also called chain A, was the first HVC discovered and has been studied in detail^{3,21}. It is a 30°-long filament, containing several well-aligned concentrations with velocities (relative to the local standard of rest, LSR) between –210 and –140 km s^{–1}. Spectra²² from the Hubble Space Telescope (HST) of the Seyfert galaxy Mark106 show strong Mg II absorption by this HVC. The fact that such absorption is not detected (it is less than 0.03 times the expected strength) in the star PG0859 + 593,

although the HVC's H I emission (as measured at 1-arcmin resolution) is similar in both directions, sets a firm lower distance limit, $d > 4 \pm 1$ kpc ($z > 2.5 \pm 0.6$ kpc), for chain A.

We have now also measured an upper limit, $d < 10$ kpc, for the upper end of chain A, using the RR Lyrae star AD Ursae Majoris, which lies at a distance of 10.1 ± 0.9 kpc (Fig. 1). Detection of interstellar lines in the spectra of RR Lyr stars is generally hampered by the presence of many stellar lines; but these are fewer during maximum phase, when the star is hotter. Figure 2 shows portions of the spectrum of AD UMa around the Ca II H and K lines, observed during maximum, and a 21-cm profile taken in the same direction. The latter has components at velocities v (relative to the LSR) of –4, –40 and –158 km s^{–1}. The Ca II K line shows the same interstellar components as the 21-cm profile, together with a strong stellar absorption around +70 km s^{–1}. The weaker Ca II H line, though blended with a broad stellar He absorption, shows similar structure. In particular, absorption by the HVC at $v \approx -160$ km s^{–1} is present at both K and H.

Could these HVC absorptions be affected by blending with stellar lines? Figure 3 shows three Fe I lines of multiplet number 4, indicating a stellar radial velocity $v = +77 \pm 2$ km s^{–1}. Comparison of their strengths and widths with those²³ in the blue field-horizontal-branch star HD161817 allows calculation (Fig. 4) of the profile of a fourth line (laboratory wavelength $\lambda_0 = 3,930.3$ Å), predicted to be present at 3,931.3 Å. Subtraction of this predicted line from the observed spectrum (Fig. 4) leaves a narrow absorption at $v = -158.2 \pm 1.2$ km s^{–1}, in close agreement with the HVC velocity $v = -157.6 \pm 0.2$ km s^{–1} found at 21 cm (Fig. 2). If this absorption were due to the Fe I 3930.3-Å line, it would have $v = +96 \pm 1$ km s^{–1} in that frame, which is clearly incompatible with the stellar velocity of $+77 \pm 2$ km s^{–1} found from the other lines in the same multiplet. The velocity difference of 19 ± 3 km s^{–1}, the lack of any other suitable identification, and the agreement with the 21-cm velocity, convincingly show that the deep line at 3,931.5 Å in Figs 3 and 4 must be due to Ca II K absorption at –158 km s^{–1} by the HVC, while the wing on the short-wavelength side is due to the stellar Fe I line.

The agreement in velocity of the high-velocity absorptions at K and H, and the fact that the ratio of line depths is about 2:1, as expected, further strengthens the identification of the HVC

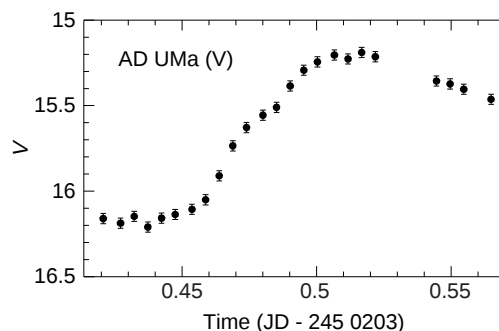


Figure 1 Partial lightcurve of the RR Lyr star AD UMa, measured in yellow light. Data were obtained on 29/30 April 1996 with the Jacobus Kapteyn Telescope at the Roque de los Muchachos Observatory by R.F.P., H.v.W. and D. Sprayberry. Minimum magnitude, 16.17 ± 0.01 ; maximum, 15.21 ± 0.01 ; average, 15.69 ± 0.01 mag. Assuming $[Fe/H] = -1.7$ by analogy with²² HD161817, recent calibrations^{33,34} of RR Lyr variables based on Hipparcos and other data yield an absolute magnitude $M_v = 0.58 \pm 0.18$. Taking an extinction³⁵ of 0.1 mag, the distance of AD UMa implied is 10.1 ± 0.9 kpc. The coordinates ($\alpha = 09^h 23^m 38.7^s$, $\delta = +55^\circ 46' 33''$ (J2000); $l = 160.40^\circ$, $b = +43.28^\circ$) were measured from the Palomar Sky Survey, with reference to the chart in the discovery paper³⁶; they differ considerably from those in the Moscow General Catalogue of Variable Stars.

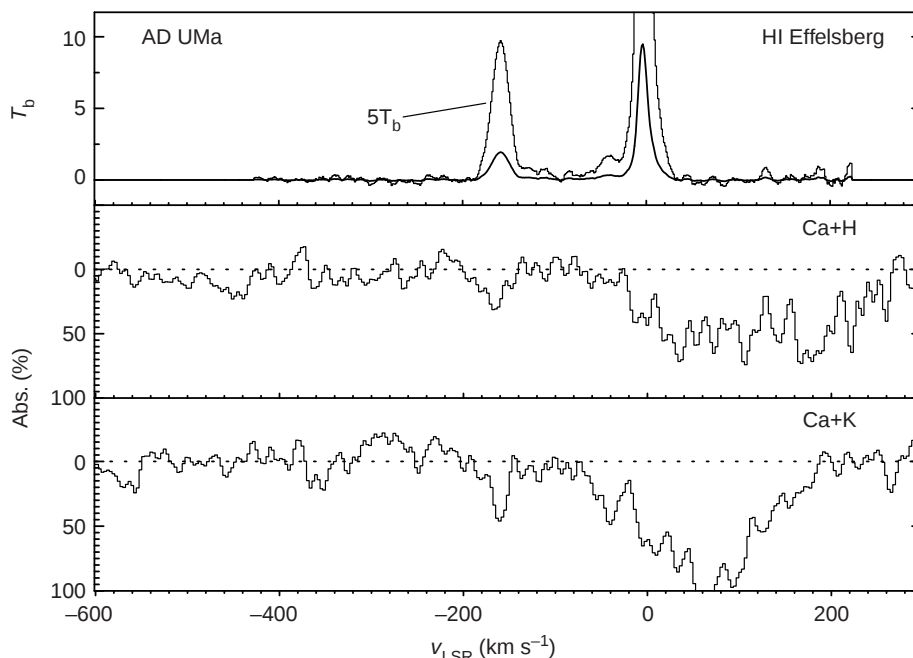


Figure 2 H I emission (top panel) and Ca II absorption measured towards AD UMa, which lies at 10 kpc distance, projected on HVC chain A. The Effelsberg 21-cm profile (top), measured at 9' beamwidth and 2 km s⁻¹ velocity resolution, is shown on two intensity scales, different by a factor 5, and has three components, at velocities (relative to the local standard of rest, LSR) of -4, -40 and -158 km s⁻¹; the last one is due to the HVC. The Ca II H and K lines, measured with the William

Herschel Telescope at La Palma at 10 km s⁻¹ resolution during maximum phase on 18 January 1997, both show—in addition to a strong stellar component near +70 km s⁻¹, and broad He absorption on the long-wavelength side of the Ca II H-line—the same interstellar components as the 21-cm line. This proves that HVC chain A, as well as the other two clouds, lies in front of the star.

absorption. Thus, it is certain that chain A lies in front of AD UMa. The Ca II and H I line strengths indicate a Ca⁺ abundance of the order of 0.01 times the total solar Ca abundance, confirming our earlier¹¹ tentative result. (Note that interstellar calcium is generally strongly depleted by inclusion into dust grains, and Ca⁺ is not the dominant ion in the interstellar gas phase.)

The absorption seen in AD UMa sets an upper limit of 10 ± 1 kpc to the distance of chain A. Combining this with the lower limit²² of 4 ± 1 kpc, we conclude that the high-latitude end of chain A lies at $4 < d < 10$ kpc, or $2.5 < z < 7$ kpc above the Galactic plane. Using the H I flux²¹, we derive an H I mass of $9,800 d^2 M_{\odot}$ (where $M_{\odot}$ is the solar mass), that is, $(1.5-10) \times 10^5 M_{\odot}$. With the standard helium

abundance, the (H I + He) gas mass would be a factor of 1.4 higher. The non-detection²⁴ of CO emission from bright H I cores in chain A implies that the H₂ mass must be one or more orders of magnitude less. Recent H α observations²⁵ suggest that chain A is mostly neutral. Assuming the ionized contribution to be minor, the total gas mass in chain A lies between 2×10^5 and $2 \times 10^6 M_{\odot}$. The kinetic energy of the complex then is of the order of $(0.3-3) \times 10^{53}$ erg, if we assume²¹ a peculiar velocity, v_{dev} , of -130 km s⁻¹.

Our distance bracket places chain A definitely in the Galactic halo, rather than in intergalactic space. It excludes models for its nature and origin requiring a distance of the order of 1 kpc or less, such as

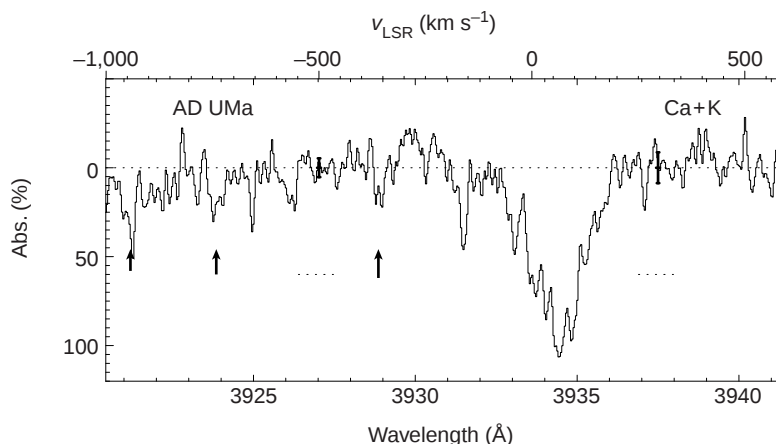


Figure 3 Spectrum of AD UMa near the Ca II K line ($\lambda_0 = 3,933.663$ Å), taken during maximum phase on 18 January 1997. The velocity scale is relative to the LSR. Note the stellar K line near +70 km s⁻¹, and the HVC absorption at -160 km s⁻¹. Three stellar Fe I lines, with $\lambda_0 = 3,920.260$, 3,922.914 and 3,927.922 Å, and shifted by

about +1.0 Å, are indicated by arrows. Gaussian fits yield velocities $v = +76$, +74 and +81 km s⁻¹. The average stellar velocity, $+77.0 \pm 2.0$ km s⁻¹, is consistent with values measured for the Fe II(4) line at 3,860 Å and, more coarsely, for the Ca II K and H δ lines. For the Fe I line with $\lambda_0 = 3,930.299$ Å, see Fig. 4.

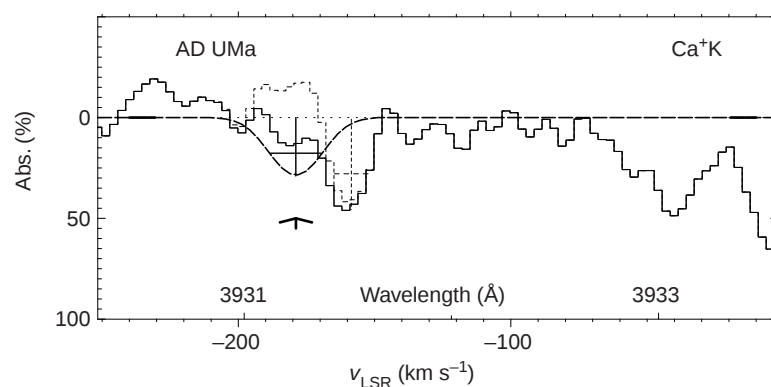


Figure 4 The predicted profile of the stellar Fe I line (with $\lambda_0 = 3,930.299 \text{ \AA}$) compared with the observed spectrum. The predicted profile is shown by a cross and a dashed line. The observed spectrum (histogram; velocity scale given for Ca II K line), in which the continuum is clearly poorly defined, barely allows an Fe I line of the predicted strength. Hence, subtraction of the predicted stellar line from the observed spectrum, leaving the short-dashed histogram spectrum, sets a lower limit to the strength of the neighbouring absorption line. This deep, narrow absorption around velocity -158 km s^{-1} (relative to LSR), fitted with a gaussian of 6 km s^{-1} dispersion (dashed cross), must be due to Ca^+ ions in HVC chain A, because it is shifted by $+19 \pm 3 \text{ km s}^{-1}$ from the predicted Fe I line. In calculating

relationships to local molecular clouds²⁶, or collision of an intergalactic cloud with the Galactic disk²⁷. It also rules out chain A being a Galactic satellite at $\sim 50 \text{ kpc}$ distance⁹, or a protogalactic gas cloud at $\sim 500 \text{ kpc}$ distance²⁸, or “a member of the Local Group of galaxies”, as proposed recently⁷ for HVCs in general. Other HVCs may well be at such great distances and fit the latter model; and some, such as the tiny, nearby cloud HVC100-7+100 (see above), may have a local origin¹⁵.

The location of chain A in the Galactic halo still allows several models for its origin. For its height $2.5 < z < 7 \text{ kpc}$ to be consistent with a Galactic-fountain model^{4,5}, a sufficiently hot halo ($T > 5 \times 10^5 \text{ K}$) would be required. The small-scale structure observed^{29,30} in chain A would then be due to instabilities formed in the downward flow of cooling clouds. Alternatively, chain A may represent gas captured from intergalactic space^{2,6,8}. In that case, collision with an ionized halo extending to high z may have served to decelerate the gas to its present velocity^{2,6,31}, and to form the small-scale structure, which has typical timescales of the order of 10^7 years (refs 3, 29), and therefore probably formed within a few kiloparsecs of its present location. In this accretion model, the question whether the origin of chain A lies in the Magellanic system (as debris from encounters between Milky Way and Magellanic Clouds), or far away in the Local Group (as “remnant of Local Group formation”^{7,8}), remains open: location in the Galactic halo does not preclude such a distant origin.

A clue to the origin of chain A might be found in its metallicity. In a Galactic fountain, near-solar metallicities would be expected; accretion might bring in HVCs with low metallicities. For chain A, current information is limited to the observed column-density ratios $N(\text{Mg}^+)/N(\text{H I}) > 0.035$ times the solar²² value, and $N(\text{Ca}^+)/N(\text{H I}) \approx 0.01$ solar (see above); in view of possible depletion and uncertain ionization conditions, these ratios only give lower limits to total Mg and Ca abundances. The best chance of obtaining a more significant value lies in measurement of the ultraviolet Si II lines in Mark106, as sulphur is not depleted by inclusion into dust grains, and S^+ is the dominant ionization stage in neutral gas. Reliable metallicity values and further direct distance measurements of HVCs hold the key to our understanding of these objects. As the HVC phenomenon is only loosely defined, and may well include objects of very different origins, distance and metallicity measurements of various HVCs will be required. □

the predicted profile, the following data were used: (1) the stellar velocity of $+77 \text{ km s}^{-1}$ discussed in Fig. 3 legend; (2) the average velocity dispersion $\sigma = 9.3 \text{ km s}^{-1}$, found from the three Fe I(4) lines with $\lambda_0 = 3,920.260$, $3,922.914$ and $3,927.922 \text{ \AA}$ (Fig. 3); and (3) an equivalent width $W = 90 \text{ m\AA}$, derived from the W values measured by us for these three lines, together with the ratio of line strengths measured²⁵ in HD161817 for the line width $\lambda_0 = 3,930.299 \text{ \AA}$ and the other three lines. The prediction procedure is supported by the fact that the line strengths found from gaussian fits to the spectrum of AD UMa are very similar to those²⁵ in the bright field-horizontal-branch star HD161817.

Received 14 December 1998; accepted 28 May 1999.

- Muller, C. A., Oort, J. H. & Raimond, E. *C.R. Acad. Sci. Paris* **257**, 1661–1664 (1963).
- Oort, J. H. Possible interpretations of the high-velocity clouds. *Bull. Astron. Inst. Neth.* **18**, 421–438 (1966).
- Wakker, B. P. & van Woerden, H. High-velocity clouds. *Annu. Rev. Astron. Astrophys.* **35**, 217–266 (1997).
- Bregman, J. N. The Galactic Fountain of high-velocity clouds. *Astrophys. J.* **236**, 577–591 (1980).
- Houck, J. C. & Bregman, J. N. Low-temperature galactic fountains. *Astrophys. J.* **352**, 506–521 (1990).
- Oort, J. H. The formation of galaxies and the origin of the high-velocity hydrogen clouds. *Astron. Astrophys. J.* **7**, 381–404 (1970).
- Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, L. & Burton, W. B. High-velocity clouds: Remnants of Local Group formation. *Bull. Am. Astron. Soc.* **28**, 1349 (1996).
- Blitz, L., Spergel, D. N., Teuben, P. J., Hartmann, L. & Burton, W. B. High-velocity clouds: Building blocks of the Local Group. *Astrophys. J.* **514**, 818–843 (1999).
- Kerr, F. J. & Sullivan, W. T. The high-velocity hydrogen clouds considered as satellites of the Galaxy. *Astrophys. J.* **158**, 115–122 (1969).
- Verschuur, G. L. High-velocity neutral hydrogen. *Annu. Rev. Astron. Astrophys.* **13**, 257–293 (1975).
- Schwarz, U. J., Wakker, B. P. & van Woerden, H. Distance and metallicity limits of high-velocity clouds. *Astron. Astrophys.* **302**, 364–381 (1995).
- Danly, L., Albert, C. E. & Kuntz, K. D. A determination of the distance to the high-velocity cloud complex M. *Astrophys. J.* **416**, L29–L31 (1993).
- Ryans, R. S. I., Keenan, F. P., Sembach, K. R. & Davies, R. D. The distance to Complex M and the Intermediate Velocity Arch. *Mon. Not. R. Astron. Soc.* **289**, 83–96 (1997).
- Bates, B., Catney, M. G. & Keenan, F. P. High-velocity gas components towards 4 Lac. *Mon. Not. R. Astron. Soc.* **242**, 267–270 (1990).
- Stoppelenburg, P. S., Schwarz, U. J. & van Woerden, H. Westerbork HI observations of two high-velocity clouds. *Astron. Astrophys.* **338**, 200–208 (1998).
- Danly, L., Lockman, F. J., Meade, M. R. & Savage, B. D. Ultraviolet and radio observations of Milky Way halo gas. *Astrophys. J. Suppl. Ser.* **81**, 125–161 (1992).
- de Boer, K. S. *et al.* The distance to the Complex C of high-velocity halo clouds. *Astron. Astrophys.* **286**, 925–934 (1994).
- Van Woerden, H., Peletier, R. F., Schwarz, U. J., Wakker, B. P. & Kalberla, P. M. W. in *Stromlo Workshop on High-Velocity Clouds* (eds Gibson, B. K. & Putman, M. E. J.) 1–25 (ASP Conf. Ser. 166, Astron. Soc. Pacif, San Francisco, 1999).
- Tamanaha, C. M. Distance constraints to the AntiCenter high-velocity clouds. *Astrophys. J. Suppl. Ser.* **104**, 81–100 (1996).
- Wakker, B. P., van Woerden, H., de Boer, K. S. & Kalberla, P. M. W. A lower limit to the distance of HVC complex H. *Astrophys. J.* **493**, 762–774 (1998).
- Wakker, B. P. & van Woerden, H. Distribution and origin of high-velocity clouds. III. Clouds, complexes and populations. *Astron. Astrophys.* **250**, 509–532 (1991).
- Wakker, B. P. *et al.* The distance to two hydrogen clouds: the high-velocity complex A and the low-latitude intermediate-velocity arch. *Astrophys. J.* **473**, 834–848 (1996).
- Adelman, S. J., Fisher, W. A. & Hill, G. An atlas of the field horizontal branch stars HD 64488, HD 109995 and HD 161817 in the photographic region. *Publ. Dominion Astrophys. Obs. Victoria* **16**, 203–280 (1987).
- Wakker, B. P., Murphy, E. M., van Woerden, H. & Dame, T. M. A sensitive search for molecular gas in high-velocity clouds. *Astrophys. J.* **488**, 216–223 (1997).
- Tufte, S. L., Reynolds, R. J. & Haffner, L. M. WHAM observations of H α emission from high-velocity clouds in the M, A, and C complexes. *Astrophys. J.* **504**, 773–784 (1998).
- Verschuur, G. L. An association between HI concentrations within high-velocity clouds A and C and nearby molecular clouds. *Astrophys. J.* **361**, 497–510 (1990).
- Meyerdierks, H. A cloud-Galaxy collision: observation and theory. *Astron. Astrophys.* **251**, 269–275 (1991).
- Verschuur, G. L. The high-velocity cloud complexes as extragalactic objects in the Local Group. *Astrophys. J.* **156**, 771–777 (1969).
- Oort, J. H. in *Problems of Physics and the Evolution of the Universe* (ed. Mirzoyan, L.) 259–280 (Acad. Sci. of Armenian SSR, Yerevan, 1978).

30. Wakker, B. P. & Schwarz, U. J. Westerbork observations of high-velocity clouds. Discussion. *Astron. Astrophys.* **250**, 484–498 (1991).
31. Benjamin, R. A. & Danly, L. High-velocity rain: The terminal velocity model of Galactic infall. *Astrophys. J.* **481**, 764–774 (1997).
32. Adelman, S. J. Elemental abundance analyses with coadded DAO spectrograms—IV. Revision of previous analyses. *Mon. Not. R. Astron. Soc.* **235**, 749–762 (1988).
33. Chaboyer, B., Demarque, P., Kernan, P. J. & Krauss, L. M. The age of globular clusters in light of Hipparcos: resolving the age problem? *Astrophys. J.* **494**, 96–110 (1998).
34. Fernley, J. *et al.* The absolute magnitudes of RR Lyrae from HIPPARCOS parallaxes and proper motions. *Astron. Astrophys.* **330**, 515–520 (1998).
35. Lucke, P. B. The distribution of color excesses and interstellar reddening material in the solar neighborhood. *Astron. Astrophys.* **64**, 367–372 (1978).
36. Hoffmeister, C. Neuer RR Lyrae-Stern S5218 Ursae Majoris. *Astron. Nachrichten* **284**, 165–166 (1958).

Acknowledgements. The William Herschel Telescope (WHT) is operated by the Royal Greenwich Observatory, in the Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, with financial support from PPARC (UK) and NWO (NL). We thank the NFRA La Palma Programme Committee, and in particular H. Henrichs, for their support for our programme. The Effelsberg Telescope belongs to the Max Planck Institute for Radio Astronomy in Bonn. B.P.W. was partly supported by NASA through STScI, which is operated by AURA, Inc, and also thanks B. Savage for financial support and useful discussions.

Correspondence and requests for materials should be addressed to H.v.W. (e-mail: hugo@astro.rug.nl).

Pairing in dense lithium

J. B. Neaton & N. W. Ashcroft

Laboratory of Atomic and Solid State Physics and the Cornell Center for Materials Research, Cornell University, Ithaca, New York 14853-2501, USA

The light alkali metals have played an important role in the developing understanding of the electronic structure of simple metals. These systems are commonly viewed in terms of an underlying interacting electron gas permeated by periodic arrays of ions normally occupying only a small fraction of the crystal volume. The electron–ion interaction, or equivalently the pseudopotential, has been argued to be weak in such systems, reflecting the partial cancellation of nuclear attraction by the largely repulsive effects of Pauli exclusion by the core electrons. Current experiments can now achieve densities at which the core electrons substantially overlap, significantly reducing the fractional volume available to fixed numbers of valence electrons. Here we report the results of first-principles calculations^{1,2} indicating that lithium, the band structure of which is largely free-electron-like at ordinary densities, does not follow the intuitive expectations of quantum mechanics by becoming even more free-electron-like at higher densities. Instead, at high pressure its electronic structure departs radically from nearly free-electron behaviour, and its common symmetric structure (body-centred cubic, b.c.c.) becomes unstable to a pairing of the ions. Once paired, lithium possesses an even number of electrons per primitive cell which, although not sufficient, is at least necessary for insulating (or semiconducting) behaviour within one-electron band theory.

The electronic structure of lithium predicted here, as revealed for example in its optical properties, leads to behaviour extremely different from that expected from the simple metals (standard Drude conductivity extended by weak interband absorption). But, as with dense hydrogen, lithium is light enough that ionic dynamics may play a significant role in stabilization energies. Within differences of such energies we find that associated with paired structures of lithium there is either an unusual semimetallic phase, corresponding to a near zero-gap semiconductor; or a metal–insulator transition may occur, resulting in a fully gapped (semiconducting) ground state as in dense hydrogen. Competing with these two possibilities is a near-insulating tri-coordinated phase, also having only a small amount of band overlap. These predictions arise in part from an application of electronic density functional theory carried out within the local density approximation (LDA)^{3,4}. Crucial to these findings is the release of the constraint, common to similar calculations, that only Bravais lattices or close equivalents are considered. When non-Bravais lattices are permitted, a paired

orthorhombic structure, *Cmca*, is found to be energetically favoured over conventional monatomic structures for $r_s < 2.1$ (the standard experimental density corresponding to $r_s = 3.25$). Here $r_s a_0 \equiv (3V/4\pi N)^{1/3}$ for N nuclei in a volume V with a_0 the Bohr radius.

The problem at hand involves core overlap, and accordingly we perform all-electron calculations with a plane-wave basis using the projector augmented wave (PAW) method⁵. The structural stability of all phases has been verified at selected densities with two other pseudopotentials, and convergence of energies with respect to plane wave cut-off and $\mathbf{k}$ -point selection has also been carefully assessed. Hellmann–Feynman forces were systematically calculated, and the ions or nuclei were steadily relaxed to equilibrium positions; all lattice vectors were optimized. A selection of structures considered (including paired structures) and their associated enthalpies is shown in Fig. 1. For b.c.c. and face-centred cubic (f.c.c.) phases the equilibrium lattice constants we find (3.36 and 4.23 Å) are in good agreement with previous LDA results^{6–8}. Lithium is known to undergo a martensitic transformation to the close-packed δ -Sm or 9R phase when cooled to 77 K at normal densities⁹, indicating the importance of lattice dynamics to the free energy; it is correctly predicted here to be favoured over b.c.c.¹⁰, f.c.c., and hexagonal close packed (h.c.p.) at the equilibrium volume, though by an amount that is comparable to the accuracy of our calculations. At a calculated pressure of 8 GPa, however, 9R reverts to f.c.c., which in turn yields to a number of lower-coordinated structures in the vicinity of 45–48 GPa; we note that all are within dynamical energies of one another (Fig. 1). The three-fold coordinated rhombohedral A7 structure emerges with lowest enthalpy for $r_s \approx 2.35$ (~ 50 GPa),

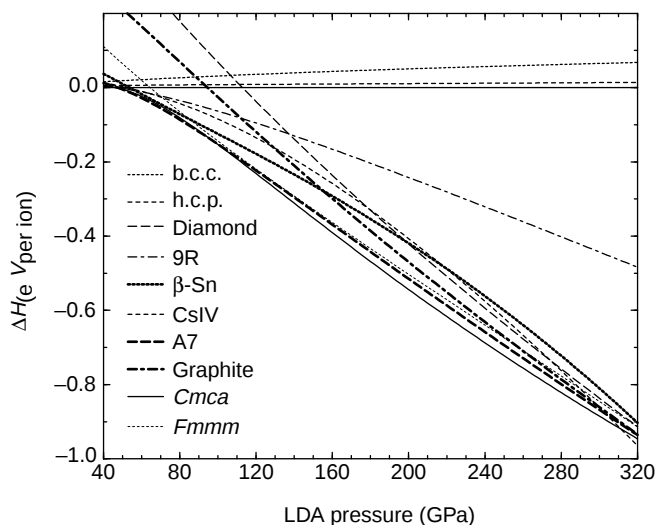


Figure 1 Enthalpy as a function of pressure for competitive structures of lithium. Enthalpy is given by $H = E + PV$; all enthalpies are referenced to the f.c.c. structure ($\Delta H = H - H_{f.c.c.}$). The total energies are well fitted to $E(V) = a_1 + a_2 V^{-2/3} + a_3 V^{-1/3} + a_4 V^{1/3} + a_5 V^{2/3}$, typical $\mathbf{k}$ -point meshes being $19 \times 13 \times 13$ for *Cmca* and $19 \times 19 \times 19$ for the monatomic structures (these result in energy convergence of 1 meV per ion). Our all electron calculations using the PAW method (having core radii of 0.58 Å) required an 80-Ry cut-off; at selected densities near $r_s = 2.1$, they were checked against a bare Coulomb potential (that is, $-3/r$ with a 220-Ry plane wave cut-off), and the A7-to-*Cmca* transition pressure (98 GPa) agreed to $\sim 1\%$. Calculations with a single-valence ultrasoft pseudopotential²¹ (incorporating the 1s states into the core) confirm most features (including the pairing) predicted with the all-electron calculations, even at densities in which there is appreciable core overlap between neighbours. The Vienna Ab-initio Simulations Package (VASP) was employed for all calculations¹²; all pseudopotentials were supplied by Kresse and Hafner²². Note that static compressions up to 40 GPa by Linn and Dunn²³ have already resulted in the observation of interesting transport properties (including increasing resistivity with pressure).

remaining the predicted ground state up to LDA pressures near 100 GPa. Taken up by group-V metals As, Sb and Bi at normal densities, A7 has a two-atom basis¹¹ and can be seen as a rhombohedral distortion of b.c.c. and simple cubic. In addition to A7, there are at least two closely competing orthorhombic paired structures at these densities, having space groups *Cmca* and *Fmmm* (the latter is obtained from *Cmca* when the dimers are oriented along [001]); there is preliminary evidence that a third structure (a derivative phase of *Pca2₁*) is also competitive in this range. Familiar from corresponding *ab initio* calculations of dense hydrogen, all have energies within dynamical corrections of A7. But when the density rises to $r_s = 2.12$ (at an LDA pressure of 98 GPa), *Cmca* has an

enthalpy lower than A7, and in this structure the lithium ions are paired, initially with a dimer distance of 1.47 Å that decreases to 1.19 Å at $r_s = 1.78$. Other competing structures at these densities are distorted 9R (six-fold coordinated); β -Sn, Cs-IV, and diamond (four-fold coordinated); and the graphite structure as well as the A7 phase (three-fold coordinated). The *Cmca* structure, in which each ion is coordinated by only one other ion, is also taken up by Ga and the halogens at normal densities. It has been predicted as the structure hydrogen first assumes in its metallic state¹².

The evident decline in coordination with increasing density has a profound effect on the nature of the electronic structure: instead of its bandwidth increasing as the density rises, we find that lithium becomes progressively less free-electron-like, culminating in the possibility of a metal–insulator transition in its paired phase for $r_s < 2.1$. As the density rises in the f.c.c. structure, the gaps at both L and X widen; the bandwidth decreases relative to these gaps, in contrast to the model behaviour expected for an alkali metal (that is, a weakly-perturbed free-electron gas in a compensating positive background). After the transition to A7, the bands depart significantly from the free-electron picture; in particular, the electronic structure is semimetallic for $r_s < 2.35$, having a nearly full lower band. As the density rises to near $r_s = 2.1$, the electronic structure of *Cmca* is also semimetallic, although it is close to that of a zero-gap semiconductor (Fig. 2). If the centres of mass of the lithium pairs are moved slightly off-site (a pairing of pairs, resulting in *Cmc2₁*) the band degeneracies in this phase are lifted, and a fully-gapped semiconducting (or insulating) phase emerges. This is identical to the behaviour observed in dense hydrogen¹²: a shift of the centres of mass (and space group, from *Cmca* to *Cmc2₁*) of the pairs opens a gap, lowering the energy of the single-particle states. The decrease in band-structure energy compensates in part for energetic cost associated with increasing electronic overlap; overall, however, any offset raises the total energy, though again it is by an amount small compared with dynamical energies, estimated to be of the

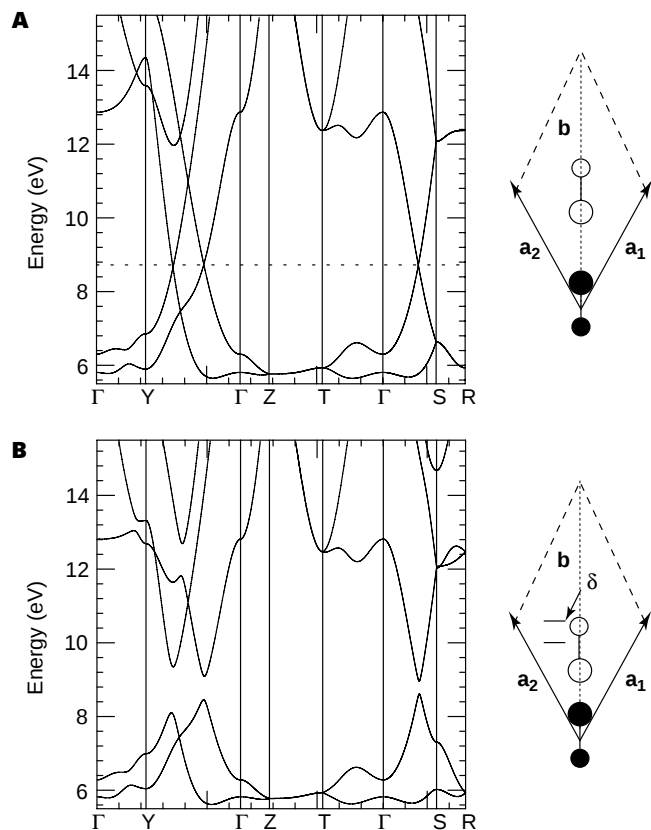


Figure 2 Band structures of dense lithium (symmetry points as given in ref. 24). **A**, *Cmca* at $r_s = 2.0$ (the dotted line is the Fermi energy) and **B**, *Cmc2₁* at the same density (with a 0.5 a.u. offset (δ) of the centre of masses of the pairs from their *Cmca* positions). The atoms are shown in the *a*-*b* plane with nearest neighbours connected with a solid line (the vertical dashed line passing through the pairs represents the *b*-axis of the non-primitive orthorhombic unit cell); the larger atoms are tilted out of the plane, the centre of mass of the atoms shown in black are in the plane, and the centre of mass of the atoms shown white are half of a lattice vector out of the plane. The lattice vectors of the primitive unit cell are labelled as $\mathbf{a}_1$ and $\mathbf{a}_2$. The offset corresponds to pairing the pairs, and opens up a direct bandgap of 0.34 eV; the insulating *Cmc2₁* phase is higher in enthalpy but by only 11 meV per ion, an energy difference small compared with the sizeable vibron ($\sim 800 \text{ cm}^{-1}$ at $r_s = 2.1$ but rising to nearly $1,350 \text{ cm}^{-1}$ at $r_s = 1.8$). Our calculated frequencies include the effects of distortions of the 1s bands, and that the 1s bandwidth (not pictured) is nearly 4 eV at this density (that is, actually wider than the 2s bands). In addition, the vibron drifts toward lower frequencies as the pairs move off-site (J.B.N. and N.W.A., manuscript in preparation), indicating that it can be important in identifying the preferred structure. The electronic structure will also have significant consequences for optical experiments: the zero-gap phase should be strongly absorptive²⁵, though pressure gradients and associated strains are expected to lead to modest band overlaps and therefore small Drude contributions; the *Cmc2₁* phase may also be highly absorptive if the overall gap associated with the displacements away from *Cmca* is quite small.

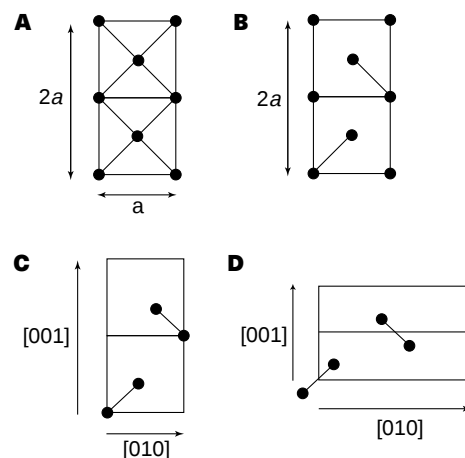


Figure 3 Two f.c.c. cells can be continuously taken into a *Cmca* structure by a Peierls-type pairing distortion. In **A** and **B**, a (100) plane is doubled along the [001] axis and as indicated the pairs are isolated. In **C** and **D** the pairing process is illustrated, leading to a lattice plane of *Cmca*. Once the ions are paired, there is significant relaxation of lattice vectors and further optimization of ionic positions as the *Cmca* moves toward its ground state. In the density range $2.4 > r_s > 2.1$, the planes perpendicular to the *c*-axis are almost evenly spaced with $c/a > b/a$, and the ions have one nearest-neighbour (although the second neighbour shell lies only slightly further than the first). Near $r_s = 2.1$, $c/a \approx b/a$ and the eight-atom orthorhombic cell is nearly tetragonal. For densities greater than $r_s = 2.1$ the ions are paired (the second neighbours lying $>20\%$ further away than the nearest), with $b/a > c/a$, $c/a \approx 1.65$ (as in dense hydrogen), and $b/a \approx 1.95$ (significantly different from $\sqrt{3}$). As density decreases f.c.c. is continuously regained if $c = 2a$, $a = b$, $d = a/\sqrt{2}$, and $\theta = 45^\circ$ (where d is the dimer length and θ is the angle with respect to the *a*-*b* (basal) plane).

order of 40 meV per ion at $r_s = 2.1$ (J.B.N. and N.W.A., manuscript in preparation). Thus the eventual structural choice will probably be determined by treatment of electrons and nuclei on an equal footing, as in hydrogen^{13,14}. For $r_s < 1.78$ ($p \approx 310$ GPa), our calculations predict that lithium finally reverts to the monatomic structure Cs-IV, analogous to a *Cmca*-to-Cs-IV transition predicted for dense hydrogen¹². Gradient corrections⁷ shift the transitions to lower-coordinated structures downward to ~ 40 GPa, and the A7-to-*Cmca* transition upward to ~ 120 GPa.

To understand the physical origin of pairing, we note that because of the increasing importance of orthogonality and exclusion with rising density, the magnitude of the pseudopotential (and its primary Fourier components) also rises. In general, any monatomic system with a strong pseudopotential is susceptible to a Peierls pairing distortion, provided that the energy penalty (electrostatic and short-range) for the pairing can be recouped by sufficient lowering of the band-structure energy. Because the magnitude of the energy drop is tied to the pseudopotential itself, its rising strength under compression in lithium suggests a critical density at which a pairing distortion will be favoured. Figure 3 shows a simple route for taking two cells of an f.c.c. structure into a single *Cmca* cell via pairing; we note that further energy lowering is achieved by reorientation of the pairs.

In a highly symmetric environment such as f.c.c., the initial pairing distortion will split the (lowest-lying) doubly degenerate bands at W; by continuity the entire band will drop, and the energy gain can in principle offset the distortion penalty. This familiar Jahn–Teller argument, supporting the general possibility of symmetry-lowering distortions in quantum-mechanical systems, complements the Peierls discussion above. The exclusionary effects imposed by the cores on the valence electrons result in an increasing departure from uniformity of the latter, and this is reflected in the band structure. We point out the remarkable flattening of bands near Γ (Fig. 4), and also note that the gap at L is nearly five times the occupied bandwidth; in addition, the 1s core bandwidth is nearly equal to the occupied valence bandwidth at this density. Finally, in

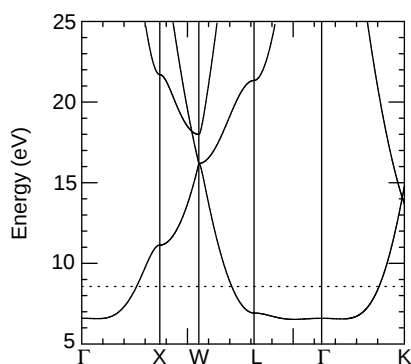


Figure 4 Band structure of f.c.c. lithium at $r_s = 2.0$. The bands are far from free-electron-like: the occupied bandwidth is < 2 eV, the Fermi surface is open at L, and the gap at L (approximately proportional to the strength of the pseudopotential) is ~ 14 eV; a two-fold degeneracy is recognized at W. The 1s bandwidth, a measure of the overlap between cores, is > 2 eV at this density. A gradual change from s-to-p character of the bands has been proposed^{26,27} above 3 Mbar, considerably higher than the onset of the pairing (~ 1 Mbar) found here. Although a change in the partial charge character may be taking place, the pairing seems not to be directly associated with it. We note that as the density rises there is considerable loss of Fermi surface in the f.c.c. phase, and this can substantially affect the optical properties. If m^* is the effective band mass and m_o the optical mass, a rigorous inequality holds that $m_o/m^* \geq (S_F/S)^2$, where S_F is the free electron Fermi surface and S is the actual (reduced) area²⁸. Because of the increasing band mass ($> 6m_e$ here, where m_e is the mass of an electron) and the decreasing Fermi surface area (evidenced by zone contact at L), the optical mass will grow with density, reducing the reflectivity.

the paired state, the disposition of electron density is itself unusual, as can be seen in Fig. 5. The exclusionary effects of the paired ions have led to regions of electron density with the appearance of connected pairs of electrons; among other possibilities, this suggests that additional exchange energy may be acquired (as in hydrogen) by the high interstitial densities. In this context, Overhauser¹⁵ has predicted charge density wave instabilities in sodium and potassium at normal densities resulting largely from exchange. For lithium, the results of spin-polarized calculations reveal spatial variations in spin density, indicating that spin wave effects may also play a role.

In the paired *Cmca* phase, the unusual character of the electronic structure may lead to conditions favouring the onset of a superconducting state: if band overlap occurs, correlated fluctuations between electron and hole bands will weaken the direct electron–electron interaction¹⁶. Plasmons can also contribute to pairing states in the valence bands, and such collective modes may also be anticipated in the unusually wide 1s bands at these densities. The possibility of superconductivity competing with the Peierls transitions cannot be ruled out in the monatomic phase: the large electron–phonon coupling (as evidenced by the strong valence electron–ion interaction) but standard Coulomb pseudopotential at these increased densities is similar to the behaviour of dense, monatomic hydrogen.

To summarize, we predict that at high pressures lithium undergoes a symmetry-breaking distortion (introduced within f.c.c. but clearly having wider applicability) resulting in a paired ground state. The initial distortion is tied to the Jahn–Teller mechanism; further distortion leading to the pairing depends on the delicate balance between exclusionary, Peierls-type band-structure effects, and exchange. The similarity to dense hydrogen is compelling, suggesting the distinct possibility of pairing in the alloys of the light alkalis, and also possibly in the alkali halides (most obviously LiH); the apparent tendency of the *Cmca* structure to occur in paired systems is also worthy of further investigation. There is already strong evidence that this effect also exists in sodium (J.B.N. and N.W.A., unpublished results), and the physics presented here suggests the possibility of paired phases in other monovalent solids, provided that unoccupied *d*-states (for example) do not fall below the Fermi energy under compression. The notion that the alkalis might become insulating at higher densities has arisen in the past¹⁷, but for ionic symmetry breaking within a fixed Bravais lattice. The mechanism proposed here is fundamentally different (and indeed no such symmetry breaking has been found up to $r_s = 2.1$), and is also predicted by general pairing arguments^{18–20}. Finite-temperature extensions of this work would be of considerable interest; in particular, pairing correlations in the liquid phase might be possible if the temperature is not too great. The possibility of an entirely

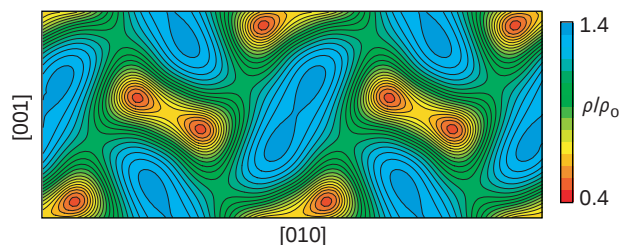


Figure 5 LDA 2s charge density in the *Cmca* structure at $r_s = 2.0$. The charge density ρ is normalized to $\rho_o = 1/(4\pi/3)r_s^3) = 0.0299$ a.u.⁻³. The ions are centred in the red regions; the ordinate points along [001] and the abscissa is along [010] of the primitive, base-centred orthorhombic cell. We note the oblate concentrations of charge in the interstitial regions (blue areas), which contain roughly two electrons (that is, a pair) and have maximum charge density ($\rho/\rho_o = 1.4$); the charge density is excluded from the regions between the ions and is lowest in the spatial regions occupied by the ions. (The 2s charge density was obtained using a single-valence pseudopotential with a 22-Ry cut-off; an equivalent plot from an all-electron calculation exhibited the same interstitial charge density.)

insulating state will probably be settled by calculations of lattice dynamical energies. □

Received 10 February; accepted 12 May 1999.

- Kresse, G. & Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B* **47**, RC558–RC561 (1993).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
- Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
- Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
- Boettger, J. C. & Trickey, S. B. Equation of state and properties of lithium. *Phys. Rev. B* **32**, 3391–3398 (1985).
- Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992).
- Jaffe, J. E. *et al.* Gaussian-basis LDA and GGA calculations for alkali-metal equations of state. *Phys. Rev. B* **57**, 11834–11837 (1998).
- Overhauser, A. W. Crystal structure of lithium at 4.2 K. *Phys. Rev. Lett.* **53**, 64–65 (1984).
- Liu, A. Y. & Cohen, M. L. Electron-phonon coupling in bcc and 9R lithium. *Phys. Rev. B* **44**, 9678–9684 (1991).
- Needs, R. J., Martin, R. M. & Nielsen, O. H. Total-energy calculations of the structural properties of the group-V element arsenic. *Phys. Rev. B* **33**, 3778–3784 (1986).
- Johnson, K. A. & Ashcroft, N. W. *Nature* (submitted).
- Straus, D. M. & Ashcroft, N. W. Self-consistent structure of metallic hydrogen. *Phys. Rev. Lett.* **38**, 415–418 (1977).
- Natoli, V., Martin, R. M. & Ceperly, D. Crystal structure of molecular hydrogen at high pressure. *Phys. Rev. Lett.* **74**, 1601–1604 (1984).
- Overhauser, A. W. Exchange and correlation instabilities of simple metals. *Phys. Rev.* **167**, 691–698 (1968).
- Richardson, C. F. & Ashcroft, N. W. High-temperature superconductivity in metallic hydrogen: electron-electron enhancements. *Phys. Rev. Lett.* **78**, 118–121 (1997).
- Siringo, F., Pucci, R. & Angilella, G. G. N. Are the light alkali metals still metals under high pressure? *High Pressure Res.* **15**, 255–262 (1997).
- Mouloupoulos, K. & Ashcroft, N. W. Scaling relations for two-component charged systems: Application to metallic hydrogen. *Phys. Rev. B* **41**, 6500–6519 (1990).
- Mouloupoulos, K. & Ashcroft, N. W. Generalized Coulomb pairing in the condensed state. *Phys. Rev. Lett.* **66**, 2915–2918 (1991).
- Mouloupoulos, K. & Ashcroft, N. W. Coulomb interactions and generalized pairing in condensed matter. *Phys. Rev. B* (in the press).
- Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892–7895 (1990).
- Kresse, G. & Hafner, J. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *J. Phys. Condens. Matter* **6**, 8245–8257 (1994).
- Lin, T. H. & Dunn, K. J. High-pressure and low-temperature study of electrical resistance of lithium. *Phys. Rev. B* **33**, 807–811 (1986).
- Cracknell, A. P. *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology* (eds Hellwege, K.-H. & Olsen, J. L.) Group III, Vol. 13c, 451–462 (Springer, Berlin, 1984).
- Sherrington, D. & Kohn, W. Frequency-dependent dielectric function of a zero-gap semiconductor. *Phys. Rev. Lett.* **21**, 153–156 (1968).
- Zittel, W. G. *et al.* Band-reordering effects in the ultra-high-pressure equation of state of lithium. *J. Phys. F* **15**, L247–L251 (1985).
- Boettger, J. C. & Albers, R. C. Structural phase stability in lithium to ultrahigh pressures. *Phys. Rev. B* **39**, 3010–3014 (1989).
- Cohen, M. H. Optical constants, heat capacity, and the Fermi surface. *Phil. Mag.* **3**, 762–775 (1958).

Acknowledgements. We thank K. A. Johnson and M. P. Teter for discussions, and we thank G. Kresse, J. Furthmüller and J. Hafner for providing the VASP software. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to N.W.A. (e-mail: nwa@ccmr.cornell.edu).

Observation of shell structure in sodium nanowires

A. I. Yanson*, I. K. Yanson*† & J. M. van Ruitenbeek*

*Kamerlingh Onnes Laboratorium, Universiteit Leiden, PO Box 9504, NL-2300 RA Leiden, The Netherlands

†B. Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences, 310164, Kharkiv, Ukraine

The quantum states of a system of particles in a finite spatial domain in general consist of a set of discrete energy eigenvalues; these are usually grouped into bunches of degenerate or close-lying levels¹, called shells. In fermionic systems, this gives rise to a local minimum in the total energy when all the states of a given shell are occupied. In particular, the closed-shell electronic configuration of the noble gases produces their exceptional stability. Shell effects have previously been observed for protons and neutrons in nuclei, and for clusters of metal atoms^{2–4}. Here we report the observation of shell effects in an open system—a

sodium metal nanowire connecting two bulk sodium metal electrodes, which are progressively pulled apart. We measure oscillations in the statistical distribution of conductance values, for contact cross-sections containing up to a hundred atoms or more. The period follows the law expected from shell-closure effects, similar to the abundance peaks at ‘magic’ numbers of atoms in metal clusters^{3,4}.

Metallic constrictions, in the form of nanowires connecting two bulk metal electrodes, have been studied down to sizes of a single atom in cross-section by means of scanning tunnelling microscopy (STM) and mechanically controllable break-junctions (MCB)^{5,6}. By indenting one electrode into another and then separating them, a stepwise decrease in electrical conductance is observed, down to the breakpoint when arriving at the last atom. Each scan of the dependence of conductance G on the elongation d is individual in detail, as the atomic configuration of each contact may be widely different. However, statistically, many scans together produce a histogram of the probability for observing a given conductance value, which is quite reproducible for a given metal and for fixed experimental parameters.

A simple description of the electronic properties of metallic nanowires is expected to work best for monovalent free-electron-like metals. The alkali metals are most suitable, as the bulk electronic states are very well described by free particles in an isotropic homogeneous positive background. In a previous experiment on sodium point contacts⁷ a histogram was obtained, which showed pronounced peaks near 1, 3, 5 and 6 times the quantum unit of conductance, $G_0 = 2e^2/h$ (where e is the charge on an electron and h is Planck’s constant). This is exactly the series of quantum numbers expected for the conductance through an ideal cylindrical conductor. We have now extended these experiments to higher temperatures and a much wider range of conductance values.

The MCB technique⁶ was adapted for the study of nanowires of the highly reactive alkali metals following ref. 7 (see Fig. 1). Scans were taken continuously by ramping the displacement d of the

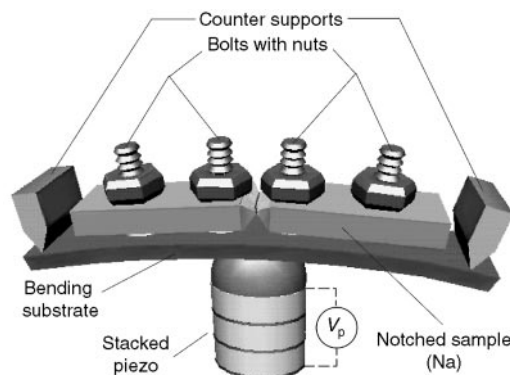


Figure 1 Schematic view of the MCB technique for alkali metals⁷. While immersed in paraffin oil, the sample is pressed onto four 1-mm-diameter brass bolts, which are glued onto the isolated substrate. Current and voltage leads are fixed with nuts onto the bolts. A notch is cut into the sample at the centre. This assembly is mounted inside a vacuum can, which is immersed in liquid helium. By bending the substrate at 4.2 K in vacuum, the sample is broken at the notch. Two fresh surfaces are exposed, and their distance is adjusted by changing the force on the substrate, employing a piezo element for fine control. The temperature of the sample is controlled by a heater and thermometer, which are fixed at the bottom of the substrate. The cryo-pumping action of the low-temperature environment ensures that the surfaces are not polluted by adsorbates. The purity of the sodium metal was at least 99.9%. The conductance was recorded using a standard d.c. voltage bias four-terminal technique, measuring the current with 16-bit resolution. The drift and calibration of the current-to-voltage converter was verified against standard resistors corresponding to 1, 10 and 100 G_0 , leading to an overall accuracy in the conductance better than 1% for $G > 10 G_0$.

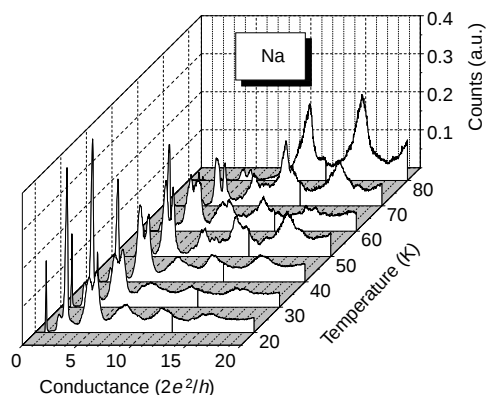


Figure 2 Temperature evolution of sodium histograms in the range from 0 to $20 G_0$. The voltage bias was 10 mV and each histogram is constructed from 1,000–2,000 individual scans. The amplitude has been normalized by the total area under each histogram (a.u., arbitrary units).

electrodes with respect to each other, using the piezo-electric driver. Each individual curve of conductance versus displacement, $G(d)$, was recorded in ~ 0.1 seconds from the highest conductance into the tunnelling regime. Histograms of conductance values were accumulated automatically involving 10^3 – 10^5 individual scans, and for a number of sample temperatures between 4.2 K and 100 K. Reproducibility and reversibility were verified by periodically returning to the same measuring conditions. Here we present results for sodium, but similar results were obtained for lithium and potassium.

Figure 2 shows the temperature dependence of the histograms for sodium in the range from 0 to $20 G_0$. In order to compare different graphs, the amplitude has been normalized by the area under each graph. The histograms at low temperatures and low conductances are similar to those given in ref. 7. We clearly recognize the familiar sharp peaks below $6 G_0$, while at higher conductances a number of rather wide maxima are found. With increasing temperature we observe a gradual decrease in amplitude of the lower-conductance peaks, while the high-conductance peaks dramatically sharpen and increase in amplitude.

Our central results are shown in Fig. 3. The histogram for sodium at temperature $T = 80$ K up to $G/G_0 = 120$ is shown in Fig. 3a inset. The smooth background shown by the dotted curve was subtracted, in order to present only the oscillating part; the latter is plotted on a semi-logarithmic scale in the main panel of Fig. 3a. Up to 17 oscillations are observed, at positions which are reproduced well for ~ 10 different samples. The influence of the procedure of background subtraction on the peak positions is much smaller than the width of the peaks.

The radius, R , of the narrowest cross-section of the nanowire can be obtained from the semi-classical expression for a ballistic wire⁸

$$G/G_0 \approx \left(\frac{k_F R}{2} \right)^2 \left(1 - \frac{2}{k_F R} \right) \quad (1)$$

where k_F is the Fermi wavevector. Using this expression, we calculate the radius of the nanowire at the peak positions (in units of k_F^{-1}); the results are shown in Fig. 3b (open squares) for consecutively numbered peak index i . Thus we observe that the peak positions are periodic as a function of the radius of the wire. This periodicity suggests a comparison of our experimental results with the magic numbers observed for sodium clusters. In the cluster experiments, a remarkable structure was observed in the distribution of cluster sizes produced in a supersonic expansion of sodium metal vapour^{3,9–13}. In the spectrum of abundance versus the number of atoms, N , in the cluster, distinct maxima are observed for $N = 2, 8, 20, 40, 58, 92, \dots$. These ‘magic numbers’ correspond to clusters having a number of valence electrons (equal to the number

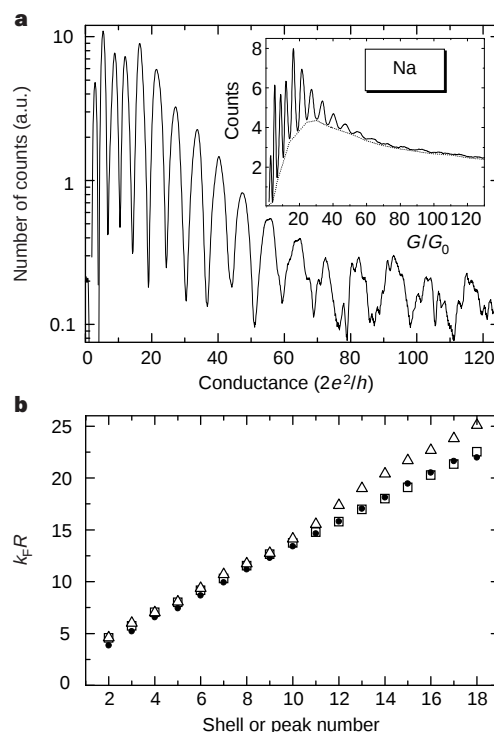


Figure 3 Conductance histograms for sodium, showing evidence for shell structure. **a**, Histogram of the number of times each conductance is observed versus the conductance in units of $G_0 = 2e^2/h$. These data are for sodium at $T = 80$ K and bias voltage $V = 100$ mV, constructed from over 10,000 individual scans. Inset, the raw data and the smooth background (dotted curve); the background is subtracted to give the curve in the main graph. The logarithmic scale for the ordinate axis helps to display the smaller-amplitude features at high conductance values. **b**, Radius of the nanowire at the positions of the maxima in **a** versus peak numbers (open squares), where R is given in units k_F^{-1} . The radii at the peak positions are compared with the radii corresponding to the magic numbers for sodium metal clusters (filled circles)¹¹ and with those expected from a semi-classical description for the fluctuations in the free energy for the nanowire (open triangles)^{18,19}.

of atoms) which just complete a shell, where the wavefunctions are considered as freely propagating waves inside a spherical potential well. The magic numbers are periodically distributed as a function of the radius of the clusters, where the radius is obtained from the number of atoms, N , in the cluster using^{3,4} $k_F R = 1.92N^{1/3}$. In Fig. 3b we compare the radii for the cluster magic numbers (filled circles) with the radii corresponding to the peaks in the conductance histogram (open squares). We establish a direct correspondence between preferential values for the conductance in histograms for sodium nanowires, and the cluster magic numbers.

The linear relation between the radii at cluster magic numbers and the shell number arises owing to fluctuations in the density of states as a function of $k_F R$, which in turn give rise to fluctuations in the free energy of the system, $\Omega(R)$. The periods of oscillation can be expressed in terms of semiclassical closed orbits inscribed inside the boundaries of the system^{1,4}, where the path length L should be an integer multiple of the Fermi wavelength $\lambda_F = 2\pi/k_F$. Fluctuations in the density of states for a free electron nanowire and their influence on the free energy has been recently considered in a number of theoretical papers^{14–19}. Using the semiclassical expansion of the free energy for a cylindrical wire^{18,19} including only the three lowest-order terms, we calculated the positions of the energy minima, which are plotted as a function of shell number in Fig. 3b (open triangles). The agreement with the experimental points is fairly good, and may be improved taking into account a more realistic shape of the potential well and higher-order contributions.

Our interpretation uses the approximately linear semiclassical relation $G(R)$ between conductance and wire cross-section, equation (1). Corrections may arise from two mechanisms. First, back scattering on defects (or phonons) may shift the peak positions. The fact that strong scattering would tend to smear the peak structure, and the close agreement between the experimental and theoretical periodicity both suggest that this shift is small. Second, the mechanism of level-bunching leading to fluctuations in $\Omega(R)$ should also give rise to fluctuation corrections to $G(R)$, which by itself would lead to peaks in the histograms even for a perfectly smooth distribution of wire diameters. Peaks due to this mechanism would be found at those points where the increase of the conductance with wire diameter is slow, and indeed we believe the peaks at low conductance (near 1, 3 and 6 G_0) are due to this mechanism. However, we argue that the main structure in Fig. 3 is due to the shell structure in $\Omega(R)$.

The main argument in favour of this interpretation comes from the temperature dependence shown in Fig. 2. As in the cluster experiments¹¹, the electronic shell structure becomes observable at higher temperatures, where the increased mobility of the atoms allows the system to explore a wider range of neighbouring atomic configurations in order to find the local energy minima. At low temperatures, the nanowires are frozen into the shape which evolves from mechanical deformations in the breaking and indenting processes. The decrease in amplitude of the sharp quantization peaks at low conductance can probably be explained by thermally induced breaking of the contact when it consists of only a few atoms. If the fluctuations in $G(R)$ were solely responsible for the observed peak structure, it could be argued that the higher temperature would favour formation of a smoother and longer wire, leading to less backscattering of electrons and less tunnelling corrections, respectively. Backscattering is responsible for the shift to lower values of the conductance peaks near 1, 3, 5 and 6 G_0 (ref. 20), and we find that the position of these peaks does not change with temperature. An indication of the length of the wire may be obtained from the global variation of the conductance with elongation²¹, and the time evolution of conductance for fixed elongation²². In our experiment, both variations show that the effective wire length decreases for increasing temperature. With these arguments, an interpretation of the temperature dependence of Fig. 2 in terms of $G(R)$ fluctuations alone is ruled out.

The correspondence between the shell structure in clusters and in nanowires can be explained by comparing the energy levels for a three-dimensional spherical potential well and a two-dimensional cylindrical geometry. The levels are obtained from the zeros of the spherical and ordinary Bessel functions, respectively. Apart from a small constant shift, the distribution of levels for the two systems has a very similar structure, with gaps and bunches of levels occurring at the same positions. This explains the striking similarity of the magic radii for the clusters and nanowires observed in Fig. 3b. □

Received 15 January; accepted 29 April 1999.

- Balian, R. & Bloch, C. Distribution of eigenfrequencies for the wave equation in a finite domain: III. Eigenfrequency density oscillations. *Ann. Phys. (NY)* **69**, 76–160 (1972).
- Bohr, Å & Mottelson, B. R. *Nuclear Structure* Vol. II (Benjamin, Reading, Massachusetts, 1975).
- de Heer, W. A. The physics of simple metal clusters: experimental aspects and simple models. *Rev. Mod. Phys.* **65**, 611–676 (1993).
- Brack, M. The physics of simple metal clusters: self-consistent jellium model and semi-classical approaches. *Rev. Mod. Phys.* **65**, 677–732 (1993).
- Agrait, N., Rodrigo, J. G. & Vieira, S. Conductance steps and quantization in atomic-size contacts. *Phys. Rev. B* **47**, 12345–12348 (1993).
- Muller, C. J., van Ruitenbeek, J. M. & de Jongh, L. J. Experimental observation of the transition from weak link to tunnel junction. *Physica C* **191**, 485–504 (1992).
- Krans, J. M. et al. The signature of conductance quantization in metallic point contacts. *Nature* **375**, 767–769 (1995).
- Torres, J. A., Pascual, J. I. & Sáenz, J. J. Theory of conduction through narrow constrictions in a three-dimensional electron gas. *Phys. Rev. B* **49**, 16581–16584 (1994).
- Knight, W. D. et al. Electronic shell structure and abundances of sodium clusters. *Phys. Rev. Lett.* **52**, 2141–2143 (1984).
- Martin, T. P. et al. Observation of electronic shells and shells of atoms in large Na clusters. *Chem. Phys. Lett.* **172**, 209–213 (1990).
- Martin, T. P. et al. Electronic shell structure of laser-warmed Na clusters. *Chem. Phys. Lett.* **186**, 53–57 (1991).

- Björnholm, S. et al. Mean-field quantization of several hundred electrons in sodium metal clusters. *Phys. Rev. Lett.* **65**, 1627–1630 (1990).
- Pedersen, J. et al. Observation of quantum supershells in clusters of sodium atoms. *Nature* **353**, 733–735 (1991).
- Stafford, C. A., Baeriswyl, D. & Bürki, J. Jellium model of metallic nanocoherence. *Phys. Rev. Lett.* **79**, 2863–2866 (1997).
- Kassubek, E., Stafford, C. A. & Grabert, H. Force, charge, and conductance of an ideal metallic nanowire. *Phys. Rev. B* **59**, 7560–7574 (1999).
- van Ruitenbeek, J. M., Devoret, M. H., Esteve, D. & Urbina, C. Conductance quantization in metals: The influence of subband formation on the relative stability of specific contact diameters. *Phys. Rev. B* **56**, 12566–12572 (1997).
- Yannouleas, C. & Landman, U. On mesoscopic forces and quantized conductance in model metallic nanowires. *J. Phys. Chem. B* **101**, 5780–5783 (1997).
- Yannouleas, C., Bogachek, E. N. & Landman, U. Energetics, forces, and quantized conductance in jellium-modeled metallic nanowires. *Phys. Rev. B* **57**, 4872–4882 (1998).
- Höppler, C. & Zwirger, W. Quantum fluctuations in the cohesive force of metallic nanowires. *Phys. Rev. B* **59**, R7849–R7851 (1999).
- Ludoph, B. et al. Evidence for saturation of channel transmission from conductance fluctuations in atomic-size point contacts. *Phys. Rev. Lett.* **82**, 1530–1533 (1999).
- Untiedt, C. et al. Fabrication and characterization of metallic nanowires. *Phys. Rev. B* **56**, 2154–2160 (1997).
- Gai, Z. et al. Spontaneous breaking of nanowires between a STM tip and the Pb(110) surface. *Phys. Rev. B* **58**, 2185–2190 (1998).

Acknowledgements. We thank L. J. de Jongh for his continuous support.

Correspondence and requests for materials should be addressed to J.M.v.R. (e-mail: Ruitenbe@Phys.LeidenUniv.nl).

Chain conformation in ultrathin polymer films

Ronald L. Jones*, Sanat K. Kumar*, Derek L. Ho†, Robert M. Briber† & Thomas P. Russell‡

* Department of Materials Science and Engineering, Pennsylvania State University, University Park, Pennsylvania 16803, USA

† Department of Materials and Nuclear Engineering, University of Maryland, College Park, Maryland 20740, USA

‡ Polymer Science and Engineering Department, University of Massachusetts, Amherst, Massachusetts 01003, USA

Polymer thin films are used in a variety of technological applications—for example, as paints, lubricants and adhesives. Theories that predict the properties of molten polymers in confined geometries (as in a thin film) generally start from the premise that the chains maintain their unperturbed gaussian conformation in the direction parallel to the surface^{1–5}. This assumption has been questioned, however, by recent experiments^{6–8}. Here we use small-angle neutron scattering to characterize the chain structure and conformation in ultrathin (less than 100 nm) polymer films. The conformation can be deduced directly from the scattering from mixtures of protonated and perdeuterated polystyrenes. We find that the gaussian conformation is retained parallel to the surfaces in all cases. Chain sizes equal the bulk value, within experimental uncertainty, although there is a systematic trend towards chain swelling in the thinnest films.

Computer simulations of polymer melts between impenetrable walls show that chain dimensions parallel to the surfaces are only slightly larger than in the bulk, and that the chain conformation in this direction remains gaussian^{9–13}. In contrast to these ideas, recent self-diffusion measurements for confined polymeric systems are consistent with the notion that chain conformation is strongly modified^{6,7}. Similarly, Frank et al.⁸ have suggested that chain structure can be significantly affected in ultrathin films (that is, thickness $D \leq 100$ nm). Preliminary scattering studies by Russell¹⁴ tentatively indicated that chain conformation is modified in the thin film geometry. Similarly, Reich and Cohen¹⁵ found that polymers of high molecular mass exhibited long-range order (up to 10 μ m) from the substrate. Based on these variable, often conflicting results, it is clear that even the most fundamental questions regarding the behaviour of polymer chains near surfaces is poorly understood.

Previously, we conducted small-angle neutron scattering (SANS) measurements¹⁶ on polymer films which were as thin as 140 nm; that is, D was ~ 10 times larger than the unperturbed radius of gyration of the chains, R_G , which were spin-cast on Si wafers. In these transmission SANS experiments the scattering vector was primarily parallel to the surface(s) of the film, and probed chain conformation in this direction. While these measurements provide an important baseline for our understanding of these situations, cases where $D \approx R_G$ are more critical for delineating the basic physics of chain conformation in the thin films. Here we present results from such cases. Since the mass of polymer in the thinnest film sample (where $D \approx 10$ nm) is four to five orders of magnitude less than a normal bulk sample, these types of thin film experiments are noise limited. In addition to the steps outlined in ref. 16, we have found that stacking several wafers (as many as 20) with identically prepared samples was necessary to improve the signal-to-noise ratio.

We investigated two isotopic polystyrene mixtures of 25 wt% perdeuterated polystyrene (dPS)/75 wt% protonated polystyrene (hPS), with matched relative molecular masses of $\sim 270,000$ ($M_w/M_n < 1.05$), or $\sim 650,000$ ($M_w/M_n < 1.05$) obtained from Polymer Source, Inc. (Dorval, Quebec, Canada). Here M_w is the weight-averaged and M_n is the number-averaged molecular weight, respectively. These chains have $R_G \sim 14$ nm and ~ 22 nm, respectively, for the low- and high-molecular-mass mixtures. These systems are expected to have upper critical solution type behaviour, with spinodal temperatures estimated to be $\sim -118^\circ\text{C}$ for the $M_n = 270,000$ (270K) mixture, and $\sim 12^\circ\text{C}$ for the 650K blend¹⁷. Samples of various thickness, ranging from 10 to 900 nm, were spin-coated from a toluene solution onto Si wafers previously cleaned with a heated acid etch (30/70 by weight $\text{H}_2\text{O}_2/\text{H}_2\text{SO}_4$) and then an HF etch (to remove the native oxide layer). The polystyrene wets the hydrogen-passivated surface which results. Film thicknesses were characterized by standard means: single-wavelength ellipsometry, profilometry and specular X-ray reflectivity. The samples were annealed at 120°C for times ranging from 8 to 100 h and quenched to room temperature into their glassy state (glass transition temperature $T_g \approx 105^\circ\text{C}$). Under these conditions the bulk blends are well mixed and far from the bulk phase boundary. The SANS measurements were conducted at the National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

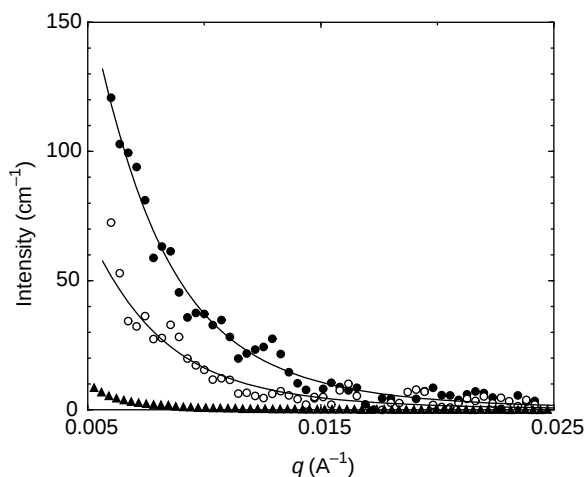


Figure 1 Plots of $I(q)$ versus q for dPS films of $M_n = 270\text{K}$. The thin films were $D = 60$ nm (open circles) and $D = 15$ nm (filled circles), respectively, while the bulk film was $D = 129\ \mu\text{m}$ (filled triangles). All films were annealed at 120°C for 12 h. SANS data were collected at NIST using a wavelength of $8\ \text{\AA}$ and a wavelength spread of 11%. The raw data were corrected for background noise, empty cell, and detector efficiency, and then placed on an absolute intensity scale relative to a silica standard and circularly averaged.

We note that, although we have tried to reduce the roughnesses of the Si surfaces, some residual roughness (~ 1 nm r.m.s. in height) persisted. In addition, there are capillary waves at the air surface of the polymer¹⁸ with an r.m.s. height of ~ 0.5 nm as measured by specular X-ray reflectivity¹⁹. To understand the role of roughness on the scattering we have cast pure dPS films of both relative molecular masses over a range of thickness. In the bulk the dPS, which has very low incoherent scattering, shows minimal scattering (Fig. 1). The excess scattering at small angles can be attributed to partial deuteration (the dPS are 99% deuterated as reported by Polymer Source)²⁰, and possibly to contaminants and voids. The thin films, on the other hand, scatter significantly, with the scattering apparently increasing strongly with decreasing thickness. Qualitatively, this scattering arises from lateral correlations of the interface roughness between the dPS and Si, as well as between dPS and air. A relatively simple explanation of these results for the thin-film dPS data can be obtained by fitting them to the Debye–Beuche equation;

$$I(q) = \frac{I_{\text{rough}}(0)}{(1 + q^2 r^2)^2} \quad (1)$$

which is reasonable to model scattering from objects with sharp interfaces. $I_{\text{rough}}(0)$ is the zero wavevector limit of the scattering function, r is the lateral correlation length of the roughness, $I(q)$ is the intensity of scattering at wavevector $q = (4\pi/\lambda)\sin\theta$, where 2θ is the scattering angle and λ is the neutron wavelength). While the general validity of this approach will be discussed in future work, it was found that all the data from the thin films could be fitted with a single value of $r = 14.8$ nm. Below, in the case of mixture films, we shall therefore model the role of roughness with equation (1) with $r = 14.8$ nm, and treat $I_{\text{rough}}(0)$ as a free parameter.

We wish to determine the contribution of concentration fluctuations to the SANS data which we shall fit to the random phase approximation (RPA) form²¹;

$$\frac{k_N}{I(q)} = \frac{1}{\phi_1 v_1 N_1 P_D(qR_G)} + \frac{1}{\phi_2 v_2 N_2 P_D(qR_G)} - 2 \frac{\chi}{v_0} \quad (2)$$

Here k_N , the neutron contrast factor, ϕ_i , the volume fraction of component i , N_i , the chain length of species i , and v_i the molar volume of i , are independently determined quantities. v_0 is a reference volume, and $P_D(x)$ is the Debye function²¹. The two

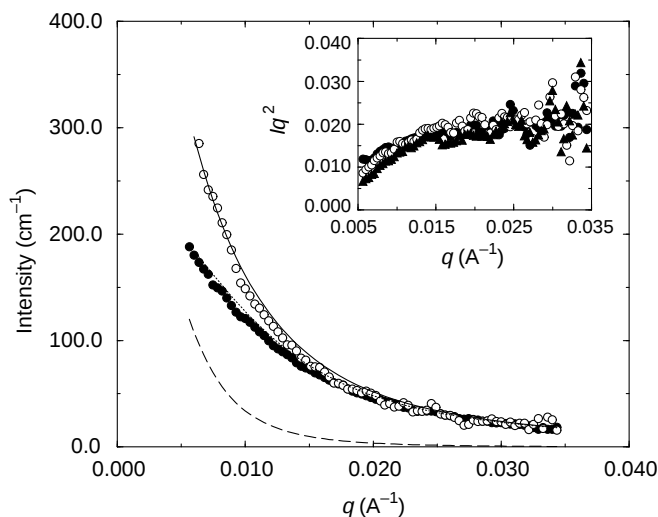


Figure 2 Plots of $I(q)$ as a function of q for dPS/hPS blends of $M_n = 270\text{K}$. Filled symbols, bulk sample ($D = 129\ \mu\text{m}$); open symbols, data for a film of $D = 18$ nm. The fit to this data (solid line) was obtained by utilizing both a roughness term (dashed line) as well as the standard RPA form (dotted line). Inset, a Kratky plot of $q^2 I(q)$ versus q for a bulk sample and several thin films of the same material, ranging from $D = 18$ nm (filled triangles) to $D = 40$ nm (filled circles).

unknowns R_G and χ , the Flory interaction parameter for this blend²², can be obtained from fits to the experimental data.

The data in Fig. 1 suggest that the contribution of surface roughness to the scattering data from the thin films manifests itself at low q values. To obtain an estimate of chain conformation without resorting to curve resolution to separate the concentration fluctuations from roughness effects, we consider data in the high- q limit. In this case, the roughness contributes minimally and equation (2) simplifies to $q^2 I(q) = 12k_N v \phi (1 - \phi) \sigma^{-2}$, where σ is the Kuhn segment length of the polymer chains in question. In Fig. 2 we plot data as $I(q)$ versus q . The inset to Fig. 2 is a Kratky plot (that is, $q^2 I(q)$ versus q) where data are shown from the 270K blends ranging from the bulk to the thinnest case of $D = 18$ nm. We have annealed the samples for various times ranging from 8 to 100 h, and found no dependence of the scattering data on annealing. Using bulk diffusion coefficients, we estimate that chains will diffuse a distance comparable to their R_G in 2.5 h for the 270K blend, and 26 h for the 650K blend. In conjunction with other results (such as annealing at 220 °C), we believe that we are not sampling non-equilibrium states in these experiments. We have not carried out a systematic investigation of shorter annealing times or for the as-spun films, where the non-equilibrium nature of the spin-casting process may be expected to affect the results. It is clear from Fig. 2 inset that the data from all films reach a q -independent plateau value, as expected. This indicates that the chains retain their gaussian conformation in a direction parallel to the surfaces, independent of film thickness. To

within the uncertainties in the data the σ values are independent of thickness, thus suggesting that the persistence length of the chains are unaffected by confinement. The value of $\sigma = 6.8$ Å for the bulk blend is in good agreement with past data¹⁷. The findings for the thin films are reasonable since $D \gg \sigma$, and hence we expect that σ should be independent of thickness. Similar results were obtained for the higher- M_n blend.

The low- q limit of the data will allow us to determine independently the R_G (which reflects chain size), and $I(0)$, which is a measure of the role of confinement on system thermodynamics. (We note that $I(0) \propto (\chi_s - \chi)^{-1}$, where χ_s is the χ value at the spinodal.) In the case of thin films we have used the form, $I(q) = I_{\text{RPA}}(q) + I_{\text{rough}}(q)$, where $I_{\text{rough}}(q)$ and $I_{\text{RPA}}(q)$ are intensity values as obtained, respectively, from equations (1) and (2), to fit the data. The lateral correlation length of the roughness was set to $r = 14.8$ nm, identical to that of the pure dPS films. In Fig. 2 we show the concentration fluctuation and roughness contributions to the SANS data from the 18-nm-thick 270K film. An excellent fit to the data is obtained using this three-parameter model.

In Fig. 3 we summarize our results for the R_G and $I(0)$ values obtained from the low- q fits for the low- and high- M_n blends, respectively. We do not show error estimates for simplicity, but the inherent variation from run to run for a given film thickness is, in all cases, comparable to the predicted systematic errors in our data. The main result is that both the R_G and $I(0)$ (or χ) values for the thin films assume essentially their bulk values for all thicknesses considered, $0.5 \leq D/R_{G,\text{bulk}} \leq \infty$, although it seems that the R_G values are always systematically larger than the bulk value for the thinnest films. These departures, however, are within the uncertainties of the data. We note that, since these mixtures are far from their bulk phase boundaries, the uncertainties in the derived χ values are large: for example, the value of χ/v_0 for the 270K blend is $(1.2 \pm 1.0) \times 10^{-6} \text{ cm}^{-3}$ independent of film thickness. It is therefore difficult for us to comment on the role of confinement on system thermodynamics. A more critical assessment of this issue could only be obtained by examining mixtures close to their phase boundaries, where the inherent uncertainties in χ will be lower.

It is very important that the interfacial roughness be properly characterized and accounted for in analysis of the experimental SANS data for films thinner than ~ 60 nm. Failure to do this can result in erroneous conclusions. For example, if the role of roughness is ignored and our thin-film data are fitted to the RPA form, then we find that, although satisfactory fits are obtained, the $I(0)$ values are strongly affected (Fig. 3); in contrast, the R_G values are effectively unchanged. So if roughness is ignored, it could be mistakenly concluded that, while chain conformation is unaffected in thin films, system thermodynamics are strongly affected by confinement. □

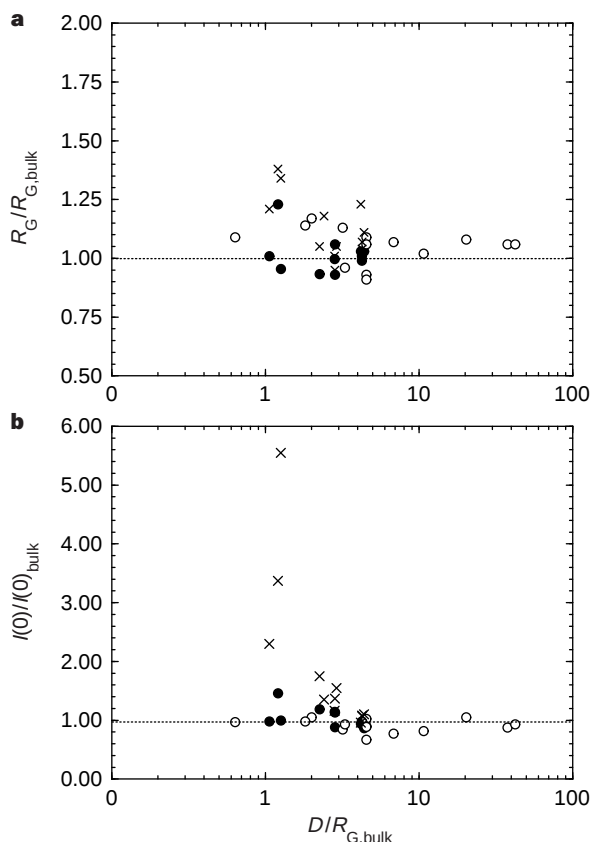


Figure 3 Plots of $I(0)/I_{\text{bulk}}(0)$ (**b**) and $R_G/R_{G,\text{bulk}}$ (**a**) as a function of $D/R_{G,\text{bulk}}$ for all of the blends studied here. Data are displayed for blends with $M_n = 270\text{K}$ (filled circles) and $M_n = 650\text{K}$ (open circles). Thin-film values of $I(0)$ and R_G are normalized by their bulk values to account for changes in instrumental configuration over the two-year period in which the data were collected. The bulk samples of the 270K and 650K materials used here were of thickness 129 μm and 163 μm , respectively. We also include data from the 270K films which were analysed without accounting for roughness, simply by fitting to the RPA form (crosses).

Received 7 December 1998; accepted 4 May 1999.

1. Fleer, G., Cohen-Stuart, M., Scheutjens, J., Cosgrove, T. & Vincent, B. *Polymers at Interfaces* (Chapman and Hall, London, 1993).
2. Balazs, A. C., Gersappe, D. & Fasolka, M. Using Monte Carlo simulations and self-consistent field theory to design interfacially active copolymers. *Macromol. Theory Simulat.* **4**, 585–612 (1995).
3. Schmidt, I. & Binder, K. Model calculations for wetting transitions in polymer mixtures. *J. Phys. (Paris)* **46**, 1631–1634 (1985).
4. Wu, D., Fredrickson, G. H. & Leibler, L. Distribution of chain ends at the surface of a polymer melt: compensation effects and surface tension. *J. Polym. Sci. B* **33**, 2373–2389 (1995).
5. Silverberg, A. Distribution of conformations and chain ends near the surface of a melt of linear flexible macromolecules. *J. Colloid Interface Sci.* **90**, 86–91 (1982).
6. Zheng, X. *et al.* Long-range effects on polymer diffusion induced by bounding surfaces. *Phys. Rev. Lett.* **79**, 241–244 (1997).
7. Calvert, P. M. When a thick film is thin. *Nature* **384**, 311–312 (1996).
8. Frank, C. W. *et al.* Structure of thin and ultrathin spin-cast polymer films. *Science* **273**, 912–915 (1996).
9. Kumar, S. K., Vacatello, M. & Yoon, D. Y. Off-lattice Monte Carlo simulations of polymer melts confined between two plates. *J. Chem. Phys.* **89**, 5206–5215 (1988).
10. Theodorou, D. N. Structure and thermodynamics of bulk homopolymer/solid interfaces: a site model approach. *Macromolecules* **21**, 1400–1410 (1988).
11. Bitsanis, I. & ten Brinke, G. A lattice Monte Carlo study of long chain conformations at solid-polymer melt interfaces. *J. Chem. Phys.* **99**, 3100–3111 (1993).
12. Pakula, T. Computer simulation of polymers in thin layers. I. Polymer melt between neutral walls—static properties. *J. Chem. Phys.* **95**, 4685–4690 (1991).
13. Yethiraj, A. & Hall, C. K. Square-well chains: Bulk equation of state using perturbation theory and

- Monte Carlo simulations of the bulk pressure and of the density profile near walls. *J. Chem. Phys.* **95**, 1999–2005 (1991).
14. Russell, T. P. *et al.* Small angle neutron scattering studies on ultrathin films. *Macromolecules* **28**, 787–789 (1995).
 15. Chen, Y. & Reich, S. Phase separation of polymer blends in thin films. *J. Polym. Sci. Polym. Phys.* **19**, 1255–1267 (1981).
 16. Ho, D. L., Briber, R. M., Jones, R. L., Kumar, S. K. & Russell, T. P. Small-angle neutron scattering studies on thin films of isotopic polystyrene blends. *Macromolecules* **31**, 9247–9252 (1998).
 17. Londono, J. D. *et al.* Composition dependence of the interaction parameter in isotopic polymer blends. *Macromolecules* **27**, 2864–2871 (1994).
 18. Rowlinson, J. S. & Widom, B. *Molecular Theory of Capillarity* (Oxford Univ. Press, New York, 1982).
 19. Russell, T. P. X-ray and neutron reflectivity for the investigation of polymers. *Mater. Sci. Rep.* **5**, 171–271 (1990).
 20. Balsara, N. P., Lohse, D. J., Graessley, W. W. & Krishnamoorti, R. Small-angle neutron scattering by partially deuterated polymers and their blends. *J. Chem. Phys.* **100**, 3905–3910 (1994).
 21. deGennes, P. G. *Scaling Concepts in Polymer Physics* (Cornell Univ. Press, Ithaca, 1979).
 22. Flory, P. J. *Principles of Polymer Chemistry* (Cornell Univ. Press, Ithaca, 1953).

Acknowledgements. This work was supported by the NSF, Division of Materials Research, the American Chemical Society (R.L.J., S.K.), the Department of Energy (T.P.R.) and the US Department of Commerce (R.M.B., D.H.). The US Department of Commerce and the NSF supported the neutron scattering facilities at NIST used in this work. We are grateful to C. Glinka for his encouragement.

Correspondence and requests for materials should be addressed to S.K.K. (e-mail: kumar@plmsc.psu.edu).

Influence of snowfall and melt timing on tree growth in subarctic Eurasia

E. A. Vaganov*, M. K. Hughes†, A. V. Kirdyanov*, F. H. Schweingruber‡ & P. P. Silkin*

* Institute of Forest SB RAS, Akademgorodok, Krasnoyarsk, 660036, Russia

† Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona 85721, USA

‡ Swiss Federal Institute of Forest, Snow and Landscape Research, Birmensdorf, Zuercherstrasse 111, CH-8903, Switzerland

The causes of a reduced sensitivity of high-latitude tree growth to variations in summer temperature for recent decades^{1,2}, compared to earlier this century, are unknown. This sensitivity change is problematic, in that relationships between tree-ring properties and temperature are widely used for reconstructing past climate. Here we report an analysis of tree-ring and climate data from the forest–tundra zone, in combination with a mechanistic model of tree-ring growth, to argue that an increasing trend of winter precipitation over the past century in many subarctic regions^{3–5} led to delayed snow melt in these permafrost environments. As a result, the initiation of cambial activity (necessary for the formation of wood cells) has been delayed relative to the pre-1960 period in the Siberian subarctic. Since the early 1960s, less of the growth season has been during what had previously been the period of

maximal growth sensitivity to temperature. This shift results not only in slower growth, but also in a reduced correlation between growth and temperature. Our results suggest that changes in winter precipitation should be considered in seeking explanations for observed changes in the timing of the ‘spring greening’ of high-latitude forests⁶, and should be taken into account in the study of the role of the Siberian subarctic forest in the global carbon cycle.

We used tree-ring measurements of conifers from the Siberian northern timberline (Fig. 1) to define relationships between tree-ring⁷ growth and temperatures at different times in the growth season (see Methods). In addition to tree-ring width measurements, maximum late-wood density was used. The most important interval of the growth season—when temperature influences cell production, and hence ring width—is quite short (Fig. 2): four pentads (periods of five contiguous days) (17 June–6 July) with significant

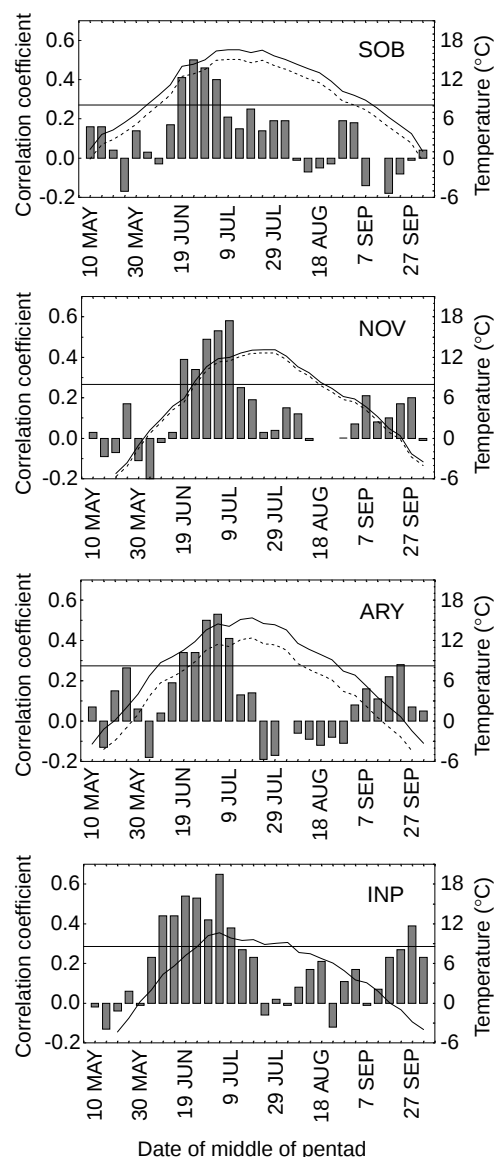


Figure 2 Correlation of the mean temperature of five consecutive days (‘pentads’) with tree-ring width indices for four sites near the northern timberline. The three-letter codes for each panel indicate the tree-ring site (see Fig. 1). At SOB, meteorological data from Berezovo were used, at NOV, from Hatanga, at ARY, from Olenek, and at INP, from Chokurdakh. Values above the horizontal solid line are statistically significant ($P < 0.05$). Solid line, mean of temperature data (over several years) from meteorological station close to site; dotted line, temperature corrected to the site.

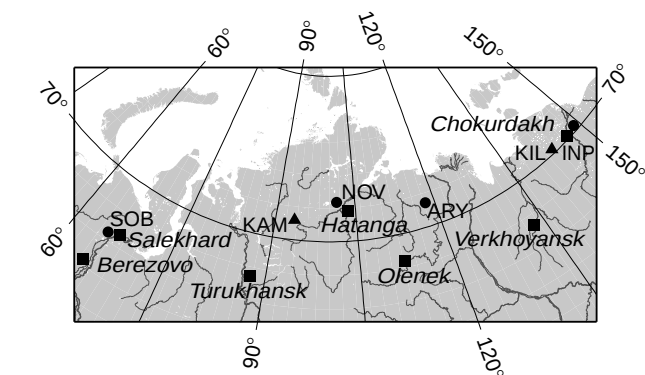


Figure 1 Location of sites and meteorological stations. Squares, meteorological stations; circles, tree-ring width sites; triangles, tree-ring width and tree-ring structure sites.

correlation at site SOB in the west, five pentads (17 June–11 July) for sites NOV and ARY, and seven pentads (7 June–6 July) for site INP in the east (station locations are shown in Fig. 1). This interval is essentially the period of the early season temperature increase. The mean temperature (averaged over several years) of the first pentad which shows significant positive correlation with tree-ring width decreases from west to east: 10 °C, 6–7 °C and 4 °C at SOB, NOV and ARY, and INP, respectively. The fraction of the season when tree-ring width and temperature are significantly correlated is much shorter than the period with temperature higher than 5 °C, increasing from 21% at SOB, to 35% at NOV and ARY, to 50% at INP. The period when temperature influences maximum late-wood density is longer; for example, 4 June–2 October at KIL.

The seasonal dynamics of temperature vary to a great extent from year to year. For example, the date after which temperature stays consistently above 0 °C can differ by up to 25–30 days. Data on the physiology of cambial activity indicate that, in conditions where temperature strongly limits the radial growth of trees, the temperature must be higher than some threshold and the thawing of the soil upper layer after snow melt must have begun so that radial growth can commence^{8–10}. Therefore, it is not strictly appropriate to correlate tree-ring parameters with temperatures for pentads whose dates are fixed relative to the calendar. Doing this will result, in some years, in tree-ring data being compared with the temperature of a period when there is no growth and, for other years, when growth is in an active, or even a very active, phase. It makes more sense to correlate tree-ring structural parameters with pentad temperatures where the dates of the pentad are fixed relative to the date of cambial initiation, as defined by tree physiology and by climate data. As soil temperature is a very important parameter for tree growth on permafrost soil¹¹, we attempted to connect the date of the beginning of growth and the date of snow melt.

The date of snow melt was calculated from temperature and winter precipitation data according to simple methods¹². The dates

calculated correlate very well with available observed snow melt data ($n = 16$, $R = 0.82$, $F = 28.8$, $P < 0.0001$) and their standard deviation, maximum and minimum values are very similar. The correlation of temperature with maximum late-wood density changed when pentad dates were fixed relative to snow melt rather than the calendar. For example, the maximum correlation between maximum late-wood density and temperature at KAM was 0.52 when calculated for calendar-related pentads, but was 0.67 when the pentads were related to the date of snow melt. This is consistent with knowledge of the physiology of cell wall growth. The variability of tree-ring width is explained by the temperature of pentads immediately after snow melt (the first part of the growth season)^{13–15}. High temperature during the first part of a season leads to formation of a wider cambial zone, higher cell production through the season, and consequently a wider ring^{16–18}. Thus, the main controlling factors of seasonal growth and tree-ring structure formation in northern timberline trees are early summer temperature and the date of snow melt, which influences the date of cambial initiation. Multiple regression models of tree-ring width indices, calculated with early summer temperature and snow-melt date as independent variables, show strong agreement with observed data. For KAM, $R^2 = 0.50$, $F = 20.3$ and $P < 0.00001$, and for KIL, $R^2 = 0.54$, $F = 22.8$ and $P < 0.00001$. Comparison of the mean ring widths of years with very early and very late snow melt confirms the influence of snow-melt date on tree-ring width. At KAM, the mean (standard deviation) tree-ring width index for the 12 years of earliest snow melt was 1.25 (0.19), and for the 10 years of latest snow melt was 0.88 (0.25). Similarly at KIL, the means (standard deviations) were 1.06 (0.28) for the 9 earliest years, and 0.92 (0.22) for the 9 latest years. These data indicate that tree-ring width indices are higher in years with earlier snow melt.

We examined the dynamics of early summer temperature, winter precipitation and calculated dates of snow melt for some meteorological stations in the study region, and for others located a

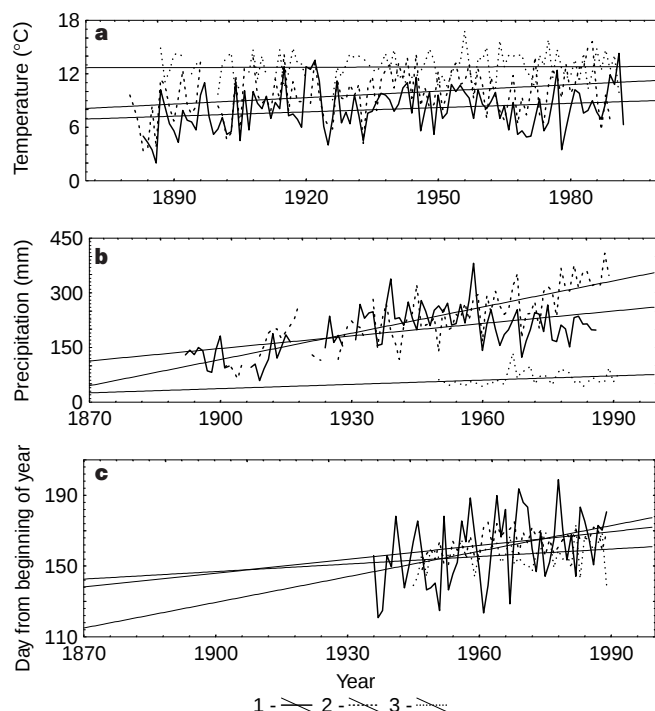


Figure 3 Meteorological time series. Variability of **a**, early summer temperature, **b**, winter precipitation and **c**, calculated dates of snow melting over several years. Data are from meteorological stations as follows: Salekhard, solid line ('1' in key), Turukhansk, broken line (2) and Verkhoyansk, dotted line (3). In each case, as well as the meteorological time series, a straight line shows the trend over the period of record. Differences in ending dates of time series relate to data availability.

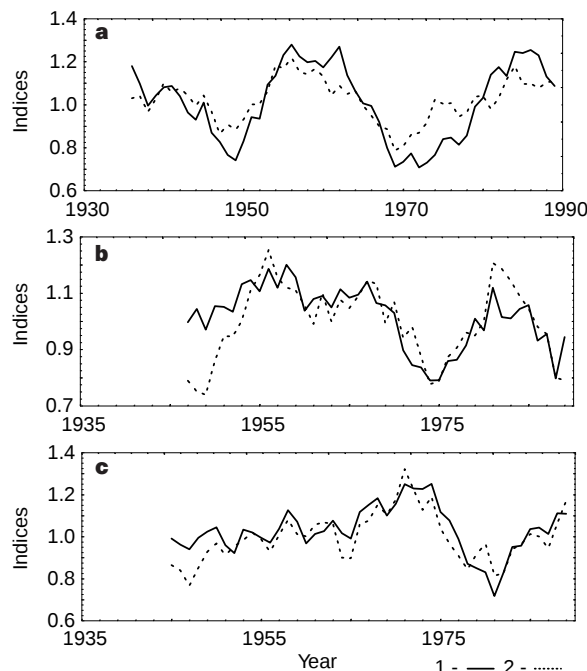


Figure 4 Tree-ring width dynamics at three sites: **(a)**, SOB; **(b)**, NOV; **(c)**, INP. Solid line ('1' in key), measurements, dashed line (2), simulation model. Data are smoothed by 5-year averaging. The vertical axis represents dimensionless indices derived by detrending the individual tree-ring width series and, in the case of the measured data, calculating a site mean.

Table 1 Parameters of trends

Region	Temperature* trend (°C per 10 yr)		Precipitation† trend (mm per 10 yr)		Date of snow melting trend (d per 10 yr)	
	Whole period	1981–90	Whole period	1981–90	Whole period	1981–90
Salekhard	+0.15	+0.13	+11.4	–36.3	+3.7	+14.1
Hatanga	+0.25	–0.70	+22.6	+22.2	+2.6	–0.3
Chokurdakh	+0.03	+0.14	+3.9	+18.0	+1.4	–3.3

* Early summer temperature.

† Winter precipitation (October–April).

short distance to the south which had a longer period of record (Fig. 3). Early summer temperature has a clear positive trend during the past century: winter precipitation also shows a positive trend. These two processes with different directions (temperature accelerates growth, winter precipitation shifts the start of cambial activity to later dates) lead to later activation of tree growth. Further, the higher the values of winter precipitation are, the greater is their influence. In the wet western and middle regions of the Siberian subarctic, the shift of snow-melt dates is larger than in the dryer eastern regions (Table 1).

To test further the combined roles of the timing of cambial initiation (linked to the date of snow melt and hence winter precipitation¹⁹) and early summer temperature in controlling tree-ring width in the subarctic, we use a simulation model^{20–22}. The model uses daily temperature, precipitation and light data to calculate seasonal tree-ring growth and cell production, taking snow melt and soil thawing into account. The model has a specific block devoted to soil thawing dynamics, so that, for example, it takes into account such possibilities as frequent refreezing and thawing. The modelled date of the first cell divisions depends on the estimated dates of the end of constant snow cover and threshold values of air temperature and effective temperature sum. Chronologies generated using this model are very similar to those collected and measured for three sites spread across the region, with correlations between modelled and observed values of 0.86 (1936–89), 0.84 (1947–89) and 0.71 (1945–89) respectively (Fig. 4). We note that these relationships do not deteriorate after 1960. This lends strong support to the hypothesis that delayed cambial activation connected with later snow melt is the real reason for observed weakening of the relationship between summer temperature and tree-ring width dynamics in the forest–tundra zone of the Siberian subarctic. There is reason to believe that these relationships may well apply through much of this vast zone, as tree-ring series are very strongly correlated over great distances in this region. In a study²³ of 65 time series derived from well replicated tree-ring widths from near the northern tree limit throughout the Siberian subarctic, over a 300-year period, pairs of tree-ring series up to 300 km apart had correlation coefficients between 0.7 and 0.8. Series from sites up to 800 km apart correlated at between 0.4 and 0.6.

Our results have implications for the study of both natural and anthropogenic variability in the subarctic. They provide an improvement of the understanding of the biological basis of summer temperature reconstructions from subarctic tree-ring variables, especially in the forest–tundra zone. Similarly, our results lead to a testable working hypothesis for the earlier and increased ‘spring greening’ of the boreal forest seen in remotely sensed primary productivity for the period 1981–90 (ref. 6). This in turn has implications for the study of the global carbon cycle, and the changes it may undergo in a changing climate. In particular, possible future changes in winter precipitation (as well as changing growth season temperatures) need to be taken into account in the development of models and climate scenarios concerning the role of the Siberian subarctic forest in the carbon cycle. □

Methods

Material from six sites near the northern timberline was analysed (Fig. 1). The

choice of these sites was defined by the presence of long records of daily temperature data. Dendroclimatic analysis of standardized tree-ring width chronologies²³ was performed for four sites. For two other sites, in addition to tree-ring width, maximum late-wood density was measured by means of X-ray microdensitometry²⁴. Local chronologies were calculated by averaging measurements made for individual trees (12–19 trees—for tree ring width, 15–20—for maximum late-wood density), and chronology statistics calculated. Correlation coefficients of the chronologies with pentad temperatures were calculated. Long instrumental records (1936–90) of daily temperature from nearby meteorological stations (Berezovo, Khatanga, Olenok and Chokurdakh) were used. In addition, monthly temperature and precipitation data available for a longer period (1887–1990) from other meteorological stations (Salekhard, Turukhansk and Verkhoyansk) were used.

Received 18 November 1998; accepted 21 May 1999.

- Briffa, K. R. *et al.* Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* **391**, 678–682 (1998).
- Briffa, K. R. *et al.* Trees tell of past climates: but are they speaking less clearly today? *Phil. Trans. R. Soc. Lond. B* **353**, 65–73 (1998).
- Findlay, B. F. *et al.* in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kiaser, D. P., Sepanski, R. J. & Stoss, F. W.) 800–828 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, Tennessee, 1994).
- Groisman, P. Ya. & Easterling, D. R. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kiaser, D. P., Sepanski, R. J. & Stoss, F. W.) 786–798 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, Tennessee, 1994).
- Groisman, P. Ya., Kokpaeva, V. V., Belokrylova, T. A. & Karl, T. R. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kiaser, D. P., Sepanski, R. J. & Stoss, F. W.) 906–910 (ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab., Oak Ridge, Tennessee, 1994).
- Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
- Cook, E. R. & Kairiukstis, L. A. (eds) *Methods of Dendrochronology: Application in Environmental Sciences* (Kluwer, Dordrecht, 1990).
- Wilson, B. F. in *The Formation of Wood in Forest Trees* (ed. Zimmermann, M. H.) 19–36 (Academic, New York, 1964).
- Kandelaki, A. A. Larch wood formation on Taymir. *Lesovedenie* **6**, 64–69 (1979). (In Russian.)
- Kandelaki, A. A. & Demyanov, V. A. Dynamics of tree ring formation of trees in Putoran mountains. *Lesovedenie* **5**, 46–49 (1982). (In Russian.)
- Pozdnyakov, L. K. *Forestry on Permafrost Soil* (Nauka, Novosibirsk, 1986). (In Russian.)
- Kuzmin, P. P. *The Snow Melting Process* (Gidrometeoizdat, Leningrad, 1961). (In Russian.)
- Zimmerman, M. H. & Brown, C. L. *Trees: Structure and Function* (Springer, Berlin, 1971).
- Gorchakovskii, P. L. & Shiyatov, S. G. *Phyto-indication of Environmental Conditions and Nature Processes in High Latitudes* (Nauka, Moscow, 1985). (In Russian.)
- Vaganov, E. A., Shashkin, A. V., Sviderskaya, I. V. & Vysotskaya, L. G. *Histometric Analysis of Woody Plant Growth* (Nauka, Novosibirsk, 1985). (In Russian.)
- Bannan, M. W. The vascular cambium and radial growth in *Thuja occidentalis* L. *Can. J. Bot.* **33**, 113–138 (1955).
- Shiyatov, S. G. *Snow Cover on the Upper Timberline and its Influence on Woody Plants* 141–157 (Inst. of Plant and Animal Ecology, Urals Branch of the Academy of Sciences USSR, Sverdlovsk, 1969). (In Russian.)
- Gregory, R. A. & Wilson, B. F. A comparison of cambial activity of white spruce in Alaska and New England. *Can. J. Bot.* **46**, 733–734 (1968).
- Groisman, P. Y., Karl, T. R. & Knight, R. W. Observed impact of snow cover on the heat balance and the rise of continental spring temperatures. *Science* **263**, 198–200 (1994).
- Vaganov, E. A., Sviderskaya, I. V. & Kondratyeva, E. N. Climatic conditions and tree ring structure: simulation model of trachidogram. *Lesovedenie* **2**, 37–45 (1990). (In Russian.)
- Fritts, H. C., Vaganov, E. A., Sviderskaya, I. V. & Shashkin, A. V. Climatic variation and tree-ring structure in conifers: a statistical simulation model of tree-ring width, number of cells, cell wall-thickness and wood density. *Clim. Res.* **1**, 37–54 (1991).
- Fritts, H. C. & Shashkin, A. V. in *Tree Rings as Indicators of Ecosystem Health* (ed. Lewis, T. E.) 17–58 (CRC Press, New York, 1995).
- Vaganov, E. A., Shiyatov, S. G. & Mazepa, V. S. *Dendroclimatic Investigation in Ural-Siberian Subarctic* (Nauka, Novosibirsk, 1996). (In Russian.)
- Lenz, O., Schär, E. & Schweingruber, F. H. Methodische Probleme bei der radiographisch-densitometrischen Bestimmung der Dichte unter der Jahrringbreiten von Holz. *Holzforschung* **30**, 114–123 (1976).

Acknowledgements. This work was supported by RFFI, CRDF, the Swiss National Science Fund, and the Earth System History Program of the US NSF.

Correspondence and requests for materials should be addressed to M.K.H. (e-mail: mhughas@lrr.arizona.edu).

Nitrogen oxide emissions after nitrogen additions in tropical forests

Sharon J. Hall* & Pamela A. Matson†

* Environmental Science Program, The Colorado College, Colorado Springs, Colorado 80903, USA

† Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305, USA

Industrial development and agricultural intensification are projected to increase in the humid tropics over the next few decades¹, increasing the emissions, transport and deposition of nitrogen-containing compounds². Most studies of the consequences of enhanced nitrogen deposition have been performed in temperate ecosystems in which biological processes are limited by nitrogen supply; they indicate that added nitrogen is retained up to decades before losses as nitrogen oxides or as nitrate (NO_3^-) begin³⁻⁵. We measured soil emissions of two gases that are important in the atmosphere, nitrous oxide (N_2O) and nitric oxide (NO), after experimental additions of nitrogen in two tropical rainforests of Hawai'i. Growth of one of the forests was limited by nitrogen; in the other, nitrogen was abundant and growth was limited by phosphorus, as is more characteristic of most tropical forests⁶. Here we show that the phosphorus-limited forest lost more nitrogen oxides than the nitrogen-limited forest, and it lost

equally large amounts after first-time additions of nitrogen as after chronic, long-term nitrogen additions. This forest seems to be naturally 'nitrogen saturated'⁷; it and perhaps other tropical forests may not retain as much anthropogenic nitrogen as do forests in northern latitudes.

Atmospheric deposition of nitrogen to terrestrial ecosystems in the Northern Hemisphere has increased dramatically in the past few decades, contributing up to 20 times background inputs to parts of the northeastern United States and Europe⁸. Chronic additions of nitrogen to these predominantly nitrogen-limited ecosystems have been shown to alter ecosystem properties and processes, including forest productivity, species composition and diversity, and losses of nitrogen from forest soil to stream water as NO_3^- and to the atmosphere as N_2O and NO (refs 5, 9–11). Many of these studies show that retention of anthropogenic nitrogen inputs in vegetation and soils is large for a period of time, sometimes up to decades, until losses increase as ecosystems approach 'nitrogen saturation', the point at which nitrogen availability exceeds biological demand.

Although nitrogen deposition networks do not currently exist in the tropics to provide empirical evidence, nitrogen deposition is also likely to be increasing there. Production and use of nitrogen fertilizer and combustion of fossil fuel are two important sources of nitrogen deposition. Over 40% of all nitrogen fertilizers are now used in the tropics and subtropics and over 60% will be used there by 2020; at the same time, fossil fuel use is expected to increase by several hundred per cent in many areas of the tropics over the coming decades².

In contrast to most temperate forests, primary production in many humid tropical ecosystems is believed to be limited by phosphorus rather than by nitrogen^{12,13}. Over 80% of tropical forests

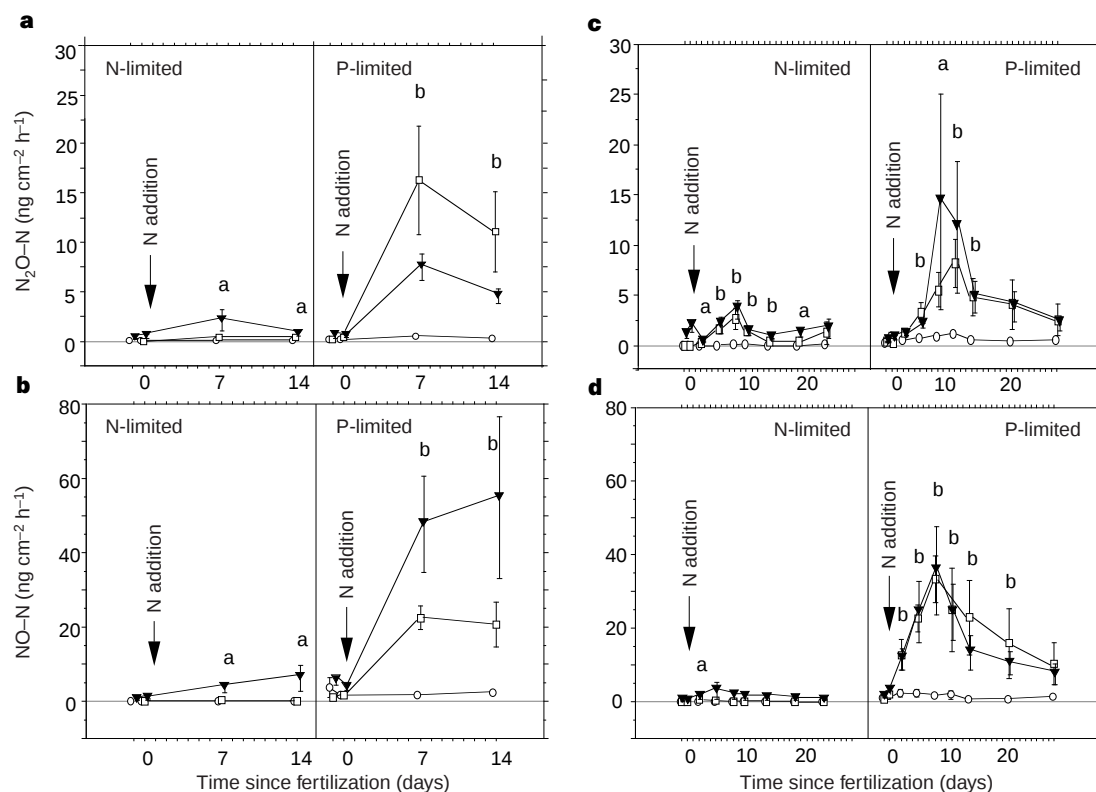


Figure 1 Nitrogen oxide emissions from soils in nitrogen (N)-limited and phosphorus (P)-limited forests before and after nitrogen addition. N_2O (**a, c**) and NO (**b, d**) emissions from soils before and immediately after nitrogen additions of 50 kg N ha^{-1} . Filled triangles, long-term fertilized (LT); open squares, first-time fertilized (FT); open circles, control (C). a, LT > C ($P < 0.05$); b, LT and FT > C ($P < 0.05$). Error bars are $\pm 1 \text{ s.e.}$ All fluxes were log transformed before ANOVA

and post-hoc Tukey analyses. **a, b**, 1996 gas emissions, 1 and 2 weeks after the first addition of nitrogen to the first-time plots. **c, d**, 1997 gas emissions after nitrogen additions; first-time plots have been fertilized for 1 yr and the long-term plots have been fertilized for 12 and 6 yr in the nitrogen-limited and phosphorus-limited sites, respectively.

Table 1 Losses of nitrogen oxide gases and gross nitrogen mineralization in the nitrogen (N)-limited and phosphorus (P)-limited forests

Site	Treatment	Fertilizer nitrogen lost as N_2O -N + NO -N in 14 days after 1996 fertilization ($kg\ N\ ha^{-1}$)	Gross nitrogen mineralization ($g\ NH_4^+ m^{-2}\ d^{-1}$)
N-limited	Control	0	0.33 (0.08) ^{a,b}
	First-time N-fertilized	0.016 (0.004) ^a	n/a
	Long-term N-fertilized	0.153 (0.049) ^b	0.58 (0.04) ^b
P-limited	Control	0	1.32 (0.13) ^c
	First-time N-fertilized	0.858 (0.187) ^b	n/a
	Long-term N-fertilized	1.366 (0.380) ^b	1.50 (0.10) ^c

* $n = 3$. Values in parentheses are ± 1 s.e. Lower-case letters represent significant differences within columns (ANOVA, $P > 0.05$).

cycle more nitrogen, and at much faster rates, than do almost all temperate forests^{13,14}, a point substantiated by recent isotopic analyses¹⁵. Furthermore, as a result of a combination of climatic factors, soil properties and rapid rates of nitrogen cycling, soils in humid tropical forests typically have higher losses of gaseous nitrogen oxides than temperate forest soils and are the largest source of N_2O globally¹⁶. N_2O is an effective greenhouse gas in the troposphere and contributes to ozone depletion in the stratosphere. NO is reactive in the troposphere and regulates regional air chemistry.

The few studies that have measured soil emissions of nitrogen oxides after one-time nitrogen additions to microplots in humid tropical forests indicate that emissions of these gases can be increased significantly^{17–19}. We expected that the temporal dynamics of nitrogen oxide losses after chronic nitrogen additions in these systems would be fundamentally different from that in temperate nitrogen-limited forests, a point that cannot be addressed using one-time fertilization experiments. Using both first-time and long-term nitrogen-addition experiments, we tested the prediction that phosphorus-limited humid tropical forests would be less able to retain anthropogenic nitrogen inputs than nitrogen-limited forests, and consequently they would not have the delayed nitrogen losses observed in well-studied nitrogen-limited systems in North America and Europe.

We measured soil N_2O and NO emissions before and after nitrogen additions in two forests located at the ends of a well defined substrate-age gradient in the Hawaiian Islands²⁰. Mean annual precipitation and temperature at both sites are 2,500 mm yr^{-1} and 16 °C (1,200 m elevation), respectively, and vary little seasonally. Both sites are dominated by the native canopy tree, *Metrosideros polymorpha*, and are on non-eroded landsurfaces with less than 6° slope. Forests at these sites have never been cleared. Geologic substrate at both sites is volcanic ash of similar chemical composition that differs in age (300 yr, soil order Inceptisol, on Hawai'i, compared with 4,100 kyr, soil order Oxisol, on Kaua'i). Long-term fertilization studies demonstrate that primary production of *M. polymorpha* is limited by nitrogen in the 300-yr site and by phosphorus in the 4,100-kyr site⁶.

In each forest, our experimental design included long-term nitrogen-addition plots (LT), 'first-time' nitrogen-addition plots (FT), and unfertilized controls (C). In the long-term treatment at the 300-yr site, nitrogen had been applied for 11 yr when our measurements began in 1996; in the 4,100-kyr site, nitrogen had been applied for 5 yr. In both sites, nitrogen (50 $kg\ N\ ha^{-1}$; 1 $ha = 10^4\ m^2$) was applied by hand twice each year to four randomly selected plots (15 m $\times$ 15 m), in the form of 50% urea and 50% ammonium nitrate (NH_4NO_3). In the first-time treatment, nitrogen was applied to previously unfertilized soil in four randomly selected plots (7.5 m $\times$ 7.5 m) at the same rate and form as the long-term treatment. Between-plot variability for each sampling date was used in ANOVA analyses ($n = 4$). N_2O and NO fluxes were measured by collecting headspace air from three closed chambers in each long-term plot, and two in each first-time plot, using methods described previously²¹. Fluxes were measured twice immediately before each nitrogen addition and between two and seven times within the one-

month period after each 50 $kg\ N\ ha^{-1}$ application. Immediately before nitrogen additions in 1996, three soil cores (0–10 cm depth) from the long-term fertilized and control plots were collected at random to determine gross nitrogen mineralization and net microbial immobilization of the nitrogen isotope ^{15}N , using standard pool dilution and ^{15}N tracer methods^{22–24}.

In the spring of 1997 we examined the sensitivity of both the nitrogen-limited and phosphorus-limited forests to anthropogenic nitrogen additions by measuring nitrogen oxide emissions after fertilization over a range of nitrogen inputs. Four groups of eight rings were installed at 5-m intervals along a transect in unfertilized forest in each site. In each group (block replicates), rings were placed more than 0.5 m apart from one another. Within each block replicate, 0, 5, 15, 25, 50, 75, 125 and 175 $kg\ N\ ha^{-1}$ was added to rings as NH_4NO_3 in 500 ml of de-ionized water by sprinkling solution within and around a 5-cm buffer outside of each ring. N_2O and NO fluxes were measured 1 day and 7 days following fertilization.

In the phosphorus-limited site, additions of nitrogen to plots fertilized for the first time caused substantial increases in both N_2O and NO during the two weeks after nitrogen additions (Fig. 1a, b). Soil nitrogen oxide emissions increased to levels typically seen in recently fertilized agricultural systems^{25,26}, and comprised $\sim 1\ kg\ N\ ha^{-1}$ of the applied nitrogen within the first two weeks (Table 1). Emissions after first-time nitrogen additions were not significantly different than those after nitrogen additions in the long-term

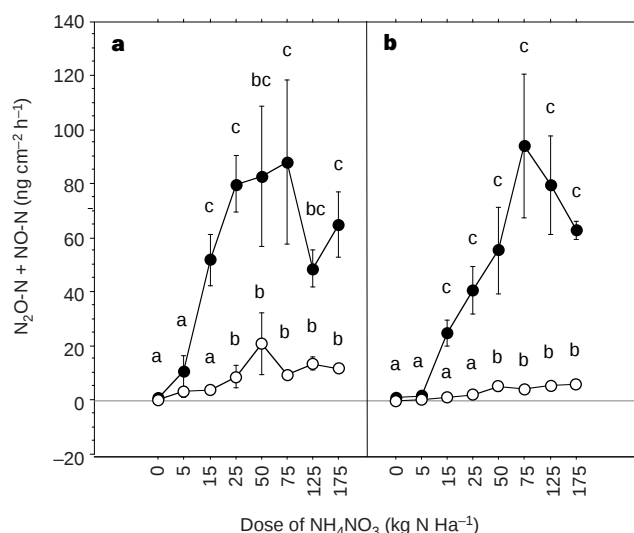


Figure 2 Nitrogen oxide emissions from soils in the N-limited and P-limited forests after a range of nitrogen additions. N_2O and NO emissions measured **a**, 1 day and **b**, 7 days after nitrogen additions in the nitrogen-limited (open circles) and phosphorus-limited (filled circles) sites. Error bars are ± 1 s.e. Error bars for the nitrogen-limited site in **b** are not visible because they are smaller than the open circles. All fluxes were log transformed before ANOVA and post-hoc Tukey analyses. Lower-case letters represent significant differences ($P < 0.05$) between treatments and sites.

Table 2 Fraction of total soil ^{15}N found in different soil pools after an 8 day incubation.

Site	Treatment*	Inorganic nitrogen				Microbial biomass nitrogen		Soil organic nitrogen	
		Initial $\text{NH}_4^+\text{-N}$ ($\mu\text{g g}^{-1}$)	Initial $\text{NO}_3^-\text{-N}$ ($\mu\text{g g}^{-1}$)	Initial APE†	Fraction of total soil ^{15}N excess (%)‡	Initial ($\mu\text{g g}^{-1}$)	Fraction of total soil ^{15}N excess (%)‡	Total N (%)	Fraction of total soil ^{15}N excess (%)‡
N-limited	C¶	6.7 (1.0)	0.3 (0.1)	9.0 (1.4)	4.5 (0.01) ^a	193.4 (4.7)	27.6 (3.5) ^a	0.9 (0.1)	67.8 (4.3) ^a
	LT	17.8 (2.9)	8.9 (6.9)	9.0 (2.5)	15.1 (2.6) ^b	218.6 (29.6)	14.0 (2.8) ^b	1.1 (0.1)	70.9 (3.2) ^a
P-limited	C	10.1 (2.1)	7.5 (0.8)	13.6 (1.5)	38.9 (6.2) ^c	265.1 (39.3)	16.6 (2.2) ^b	1.5 (0.3)	44.5 (6.3) ^{ab}
	LT	13.5 (3.5)	12.6 (3.4)	11.8 (1.0)	57.9 (10.6) ^c	292.7 (28.8)	17.8 (1.8) ^b	1.6 (0.2)	24.3 (10.1) ^b

* C, control; LT, long-term nitrogen fertilized.

† Initial APE is atom % ^{15}N excess as inorganic nitrogen measured 15 min after labelling.

‡ Lower-case letters represent significant differences within columns (ANOVA, $P < 0.05$).

¶ $n = 3$.

fertilized plots. Patterns were similar in 1997 (Fig. 1c, d). In the one-month period after nitrogen additions in 1997, 1.3 (± 0.5) kg ha^{-1} (± 1 s.e.) of applied nitrogen were lost as N_2O -N and NO -N from the first-time fertilized plots in the phosphorus-limited site, which is similar to fluxes in the long-term fertilized plots (1.2 ± 0.2 kg N ha^{-1}). Results from the dose-response experiment indicated that nitrogen additions of 15 kg N ha^{-1} or greater resulted in elevated emissions that were not significantly different from one another (Fig. 2a, b). In the 15 kg N ha^{-1} treatment, 4% of the added nitrogen was lost as N_2O and NO over the 7-day sampling period.

Emissions of nitrogen oxides from temperate forest soils are typically low²⁷ but have been shown to increase significantly in response to elevated nitrogen deposition as systems reach nitrogen saturation over time^{5,28}. Our nitrogen-limited site behaved similarly: fluxes of nitrogen oxides in controls were very low and did not increase significantly after the first nitrogen addition. Fluxes after nitrogen additions in the long-term fertilized plots were significantly greater than controls but were small compared with fluxes in the phosphorus-limited site (Fig. 1a, b; Table 1). In 1997, only 0.06 (± 0.01) and 0.15 (± 0.06) kg N ha^{-1} were lost from the first-time fertilized and long-term fertilized plots, respectively (Fig. 1c, d). Furthermore, emissions of nitrogen oxides were minimal after all doses of nitrogen in this nitrogen-limited forest: less than 0.6% of the applied nitrogen at any dosage was lost as nitrogen oxides in 7 days (Fig. 2a, b).

Differences in soil nitrogen dynamics between these two sites help to explain the differences in nitrogen oxide emissions that we observed after fertilization. Unfertilized soil in the nitrogen-limited site was an extremely strong nitrogen sink: in control soils, 68% ($\pm 9\%$) of our $^{15}\text{NH}_4^+$ label (measured as total soil ^{15}N excess) was removed from the inorganic fraction within 15 min after addition, compared with 13% ($\pm 5\%$) removed in long-term fertilized soils. In this nitrogen-limited site, rates of gross nitrogen mineralization (conversion of organically bound nitrogen to NH_4^+) increased and net microbial nitrogen immobilization (nitrogen uptake) declined significantly after long-term nitrogen additions (Table 2), indicating a change in microbial nitrogen use over time owing to nitrogen inputs. These results indicate that repeated nitrogen fertilization decreases microbial demand for nitrogen in this site as microbial growth becomes limited by some other factor. Similar findings have been reported for both experimentally fertilized and nitrogen-deposition-affected temperate forests of Europe²⁹. In contrast, inputs of inorganic nitrogen were not quickly removed in soil of our phosphorus-limited system. Fifteen minutes after labelling, 86% ($\pm 7\%$) of our $^{15}\text{NH}_4^+$ label remained in inorganic form in control soils, and this was not statistically different than the proportion that remained in the long-term fertilized soils ($97\% \pm 6\%$). Furthermore, chronic nitrogen additions did not significantly alter the already high availability of nitrogen in this phosphorus-limited system. Neither gross nitrogen mineralization nor net NH_4^+ immobilization were altered by long-term fertilization (Tables 1 and 2).

Our results suggest that nitrogen oxide emissions from phosphorus-limited tropical ecosystems may be especially sensitive to anthropogenic nitrogen inputs, in that they may respond to even small initial nitrogen additions with larger losses than predicted by models developed for temperate forests. It has been estimated that 22 Tg of nitrogen is currently deposited yearly to terrestrial ecosystems due to fossil fuel combustion², occurring primarily in the temperate zone. By year 2020, anthropogenic nitrogen deposition is estimated to increase by approximately 25% in industrialized regions and will at least double in less industrialized regions of the world². We shall assume that 22 Tg N yr^{-1} will be deposited on tropical forests 50 years from now, that tropical forests will be weak sinks for nitrogen, and that 2% of that nitrogen will be emitted to the atmosphere as N_2O and 1% as NO (assuming 4% of inputs will be lost as we found in our 15 kg N ha^{-1} dose-response experiment and 50% of the NO will be taken up by the forest canopy¹⁷). Under these assumptions, anthropogenic deposition will enhance fluxes from tropical forests by 0.4 Tg N yr^{-1} for N_2O and 0.2 Tg N yr^{-1} for NO . This represents a 13% and 18% increase over current tropical evergreen forest N_2O emissions (3 Tg N yr^{-1})¹⁶ and NO emissions (1.1 Tg N yr^{-1})²⁷, respectively, and has potentially serious implications for global atmospheric composition and regional tropospheric photochemistry. This study does not simulate atmospheric nitrogen deposition directly, and these forests are not representative of tropical forests worldwide; however, our results suggest that there is an important difference in responses to nitrogen additions between nitrogen-limited and phosphorus-limited systems, one that should be tested in other humid tropical forests near areas undergoing industrial and agricultural development. □

Methods

Measurement of gas flux. N_2O fluxes were measured in the field by sampling air from three randomly placed, 5-l closed chambers four times over a 30-min period. Air samples and certified standards were analysed for N_2O in the laboratory using a gas chromatograph equipped with a ^{63}Ni electron-capture detector. NO was measured *in situ* by using a portable chemiluminescent detector attached to a closed, 5-l chamber by Teflon tubing and calculated as the linear increase in NO concentrations within the chamber over a 4-min period. All NO was converted to NO_2 by a CrO_3 filter, and concentration of NO_2 was estimated photochemically after reaction with Luminol solution. The chemiluminescent detector was calibrated several times a day to a known NO standard in the field and further compared to another NO standard in the laboratory.

Calculations of gas loss (Table 1). Gas losses per sampling period were calculated first by summing daily N_2O and NO fluxes within each plot on each sampling date, and then summing the combined $\text{N}_2\text{O} + \text{NO}$ flux across all sampling dates, assuming a constant rate of emissions over 24 h and a linear change in emissions between sampling days. In the nitrogen-limited site, mean control fluxes were subtracted from fertilized fluxes in each plot on each sampling date. In the phosphorus-limited site where plots were blocked, control fluxes were subtracted from fertilized fluxes within each block. Fluxes were log transformed before ANOVA and Tukey analyses.

^{15}N assays. Three soil cores were collected randomly within each plot (0–10

cm depth), 6 months after the last fertilization event in 1996, combined in the field, hand mixed and hand sieved (to 4 mm) and incubated with $^{15}\text{NH}_4^+$ in polyethylene bags in the laboratory. Approximately $1.6 \mu\text{g } ^{15}\text{N g}^{-1}$ were added to soils as $(^{15}\text{NH}_4)_2\text{SO}_4$ in 3 ml of water and comprised 30% or less of the average standing pool size of $\text{NH}_4^+\text{-N}$ at each site. Gross mineralization values were log transformed before ANOVA and Tukey analyses.

^{15}N partitioning (Table 2). The quantity of labelled nitrogen in various soil pools was calculated by multiplying nitrogen pool size by the measured excess ^{15}N concentration and then dividing by the pool of total soil ^{15}N . Microbial biomass ^{15}N was estimated first by subtracting the label present in the non-fumigated K_2SO_4 -extractable fraction from the fumigated K_2SO_4 -extractable fraction, then calculating as above. The fraction found in soil organic nitrogen was calculated by subtracting the label in microbial biomass and in inorganic nitrogen ($\text{NH}_4^+\text{-N}$ and $\text{NO}_3^+\text{-N}$) from the label in total soil nitrogen.

Received 8 March; accepted 12 May 1999.

- Vitousek, P. M. & Matson, P. A. in *Biogeochemistry of Global Change: Radiatively Active Trace Gases* (ed. Oremland, R. S.) 193–208 (Chapman & Hall, London, 1993).
- Galloway, J. N. Anthropogenic mobilization of sulphur and nitrogen: immediate and delayed consequences. *Annu. Rev. Energy Environ.* **21**, 261–292 (1996).
- Magill, A. H. et al. Biogeochemical response of forest ecosystems to simulated chronic nitrogen deposition. *Ecol. Appl.* **7**, 402–415 (1997).
- Gundersen, P., Emmett, B. A., Kjonaas, O. J., Koopmans, C. J. & Tietema, A. Impact of nitrogen deposition on nitrogen cycling in forests: a synthesis of NITREX data. *For. Ecol. Mgmt* **101**, 37–55 (1998).
- Fenn, M. E., Poth, M. A. & Johnson, D. W. Evidence for nitrogen saturation in the San Bernardino Mountains in Southern California. *For. Ecol. Mgmt* **82**, 211–230 (1996).
- Vitousek, P. M. et al. Soil and ecosystem development across the Hawaiian Islands. *GSA Today* **7**, 1–8 (1997).
- Aber, J. D., Nadelhoffer, K. J., Steudler, P. & Melillo, J. M. Nitrogen saturation in northern forest ecosystems. *Bioscience* **39**, 378–386 (1989).
- Lovett, G. M. Atmospheric deposition of nutrients and pollutants in North America: an ecological perspective. *Ecol. Appl.* **4**, 629–650 (1994).
- Kauppi, P. E., Mielikainen, K. & Kuusela, K. Biomass and carbon budget of European forests, 1971–1990. *Science* **256**, 70–74 (1992).
- Dise, N. B. & Wright, R. F. Nitrogen leaching from European forests in relation to nitrogen deposition. *For. Ecol. Mgmt* **71**, 153–161 (1995).
- Berendse, F., Aerts, R. & Bobbink, R. in *Landscape Ecology of a Stressed Environment* (eds Vos, C. C. & Opdam, P.) 105–121 (Chapman & Hall, London, 1993).
- Tanner, E. V. J., Vitousek, P. M. & Cuevas, E. Experimental investigation of nutrient limitation of forest growth on wet tropical mountains. *Ecology* **79**, 10–22 (1998).
- Vitousek, P. M. & Sanford, R. L. Nutrient cycling in moist tropical forest. *Annu. Rev. Ecol. Syst.* **17**, 137–167 (1986).
- Matson, P. A. & Vitousek, P. M. Cross-system comparisons of soil nitrogen transformations and nitrous oxide flux in tropical forest ecosystems. *Global Biogeochem. Cycles* **1**, 163–170 (1987).
- Martinelli, L. A., Piccola, M. C., Townsend, A. R. & Vitousek, P. M. Nitrogen stable isotope composition of leaves and soils: tropical versus temperate forests. *Biogeochemistry* (in the press).
- Prather, M. J. et al. in *Climate Change 1994* (eds Houghton, J. T. et al.) 23–46 (Cambridge Univ. Press, 1994).
- Bakwin, P. S. et al. Emissions of nitric oxide (NO) from tropical forest soils and exchange of NO between the forest canopy and atmospheric boundary layers. *J. Geophys. Res.* **95**, 16755–16764 (1990).
- Kaplan, W. A., Wofsy, S. C., Keller, M., & DaCosta, J. M. Emission of NO and deposition of O_3 in a tropical forest system. *J. Geophys. Res.* **93**, 1389–1395 (1988).
- Keller, M., Kaplan, W. A., Wofsy, S. C., & DaCosta, J. M. Emissions of N_2O from tropical forest soils: response to fertilization with NH_4^+ , NO_3^- , and PO_4^{3-} . *J. Geophys. Res.* **93**, 1600–1604 (1988).
- Crews, T. E. et al. Changes in soil phosphorus fractions and ecosystem dynamics across a long chronosequence in Hawai'i. *Ecology* **76**, 1407–1424 (1995).
- Matson, P. A., Billow, C., Hall, S. & Zachariassen, J. Fertilization practices and soil variations control nitrogen oxide emissions from tropical sugar cane. *J. Geophys. Res.* **101**, 18533–18545 (1996).
- Davidson, E. A., Eckert, R. W., Hart, S. C. & Firestone, M. K. Direct extraction of microbial biomass nitrogen from forest and grassland soils of California. *Soil Biol. Biochem.* **21**, 773–778 (1989).
- Hart, S. C., Stark, J. M., Davidson, E. A. & Firestone, M. K. in *Methods of Soil Analysis*, Part 2 (Microbiological and Biochemical Properties) 985–1018 (Soil Science Society of America, Madison, 1994).
- Stark, J. M. & Hart, S. C. Diffusion technique for preparing salt solutions, Kjeldahl digests, and persulfate digests for nitrogen-15 analysis. *Soil Sci. Soc. Am. J.* **60**, 1846–1855 (1996).
- Eichner, M. J. Nitrous oxide emissions from fertilized soils: summary of available data. *J. Environ. Quality* **19**, 272–280 (1990).
- Hall, S. J. & Matson, P. A. NO_x emissions from soil: Implications for air quality modeling in agricultural regions. *Annu. Rev. Energy Environ.* **21**, 311–346 (1996).
- Davidson, E. A. & Kingler, W. A global inventory of nitric oxide emissions from soils. *Nut. Cycl. Agroecosys.* **48**, 37–50 (1997).
- Butterbach-Bahl, K., Gasche, R., Huber, C. H. & Kreutzer, K. Impact of N-input by wet deposition on N-trace gas fluxes and CH_4 -oxidation in spruce forest ecosystems of the temperate zone in Europe. *Atmos. Environ.* **32**, 559–564 (1998).
- Tietema, A. Microbial carbon and nitrogen dynamics in coniferous forests floor material along a European nitrogen deposition gradient. *For. Ecol. Mgmt* **101**, 29–36 (1998).

Acknowledgements. We thank Hawai'i Volcanoes National Park and Koke'e State Park for access to and use of field sites; P. Vitousek, J. Moen and H. Farrington for facilitating all aspects of the field work; T. Billow for facilitating field and lab work; D. Herman, D. Turner, P. Brooks, H. Farrington, J. Moen, T. Sprunk, M. Mack, T. Schuur, S. Hobbie, B. Ostertag, H. Pearson, J. Funk, J. Wagner and C. Young for assistance in field and lab; and P. Vitousek, C. D'Antonio, M. Firestone and F. Thornton for comments on earlier drafts of the manuscript. We acknowledge the support of the USDA Terrestrial Ecosystem and Global Change Program and the Andrew Mellon Foundation to P.A.M., and the NASA Earth Systems Science Fellowship Program and the NSF Pre-doctoral Fellowship Program to S.J.H.

Correspondence and requests for materials should be addressed to S.J.H. (e-mail: shall@coloradocollege.edu).

Single gene circles in dinoflagellate chloroplast genomes

Zhaoduo Zhang, B. R. Green & T. Cavalier-Smith

Canadian Institute for Advanced Research Evolutionary Biology Programme, Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Photosynthetic dinoflagellates are important aquatic primary producers and notorious causes of toxic 'red tides'. Typical dinoflagellate chloroplasts differ from all other plastids in having a combination of three envelope membranes¹ and peridinin-chlorophyll *a/c* light-harvesting pigments². Despite evidence of a dinoflagellate satellite DNA containing chloroplast genes³, previous attempts to obtain chloroplast gene sequences have been uniformly unsuccessful. Here we show that the dinoflagellate chloroplast DNA genome structure is unique. Complete sequences of chloroplast ribosomal RNA genes and seven chloroplast protein genes from the dinoflagellate *Heterocapsa triquetra* reveal that each is located alone on a separate minicircular chromosome: 'one gene—one circle'. The genes are the most divergent known from chloroplast genomes. Each circle has an unusual tripartite non-coding region (putative replicon origin), which is highly conserved among the nine circles through extensive gene conversion, but is very divergent between species. Several other dinoflagellate species have minicircular chloroplast genes, indicating that this type of genomic organization may have evolved in ancestral peridinin dinoflagellates. Phylogenetic analysis indicates that dinoflagellate chloroplasts are related to chromistan and red algal chloroplasts and supports their origin by secondary symbiogenesis^{4–6}.

Total DNA from axenic *H. triquetra* was separated into satellite and major DNA bands by centrifugation on CsCl gradients in the presence of an intercalating dye³. A spinach chloroplast gene probe (*psbA*) hybridized only to satellite DNA (Fig. 1, lanes 1–4). The satellite DNA was partially digested with *Sau3A*, and three plasmid

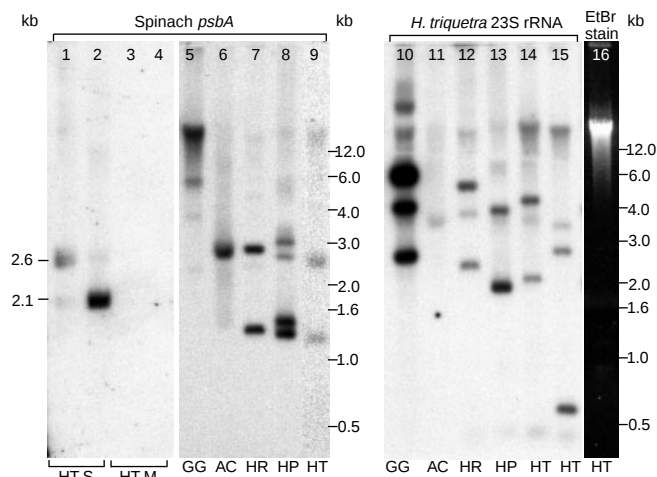


Figure 1 Hybridization of dinoflagellate DNA with chloroplast gene probes. Lanes 1, 2, *H. triquetra* satellite DNA (HT-S); 3, 4, *H. triquetra* main band DNA (HT-M); 5–16, total DNA of *H. triquetra* (HT), *H. pygmaea* (HP), *H. rotundata* (HR), *A. carterae* (AC) and *G. grindleyi* (GG). Lanes 2 and 4 were treated with $0.5 \mu\text{l}^{-1}$ *EcoR1* for 2 h; lane 15 with $0.5 \mu\text{l}^{-1}$ *BglII* for 2 h. Probes were spinach *psbA* (lanes 1–9) and HT 23S rRNA (lanes 10–15). Lane 16 is ethidium bromide stained DNA of lane 14.

Table 1 Features of nine *H. triquetra* chloroplast genes and their 9G-9A-9G region

Gene	No. of clones	Protein of	Circle (bp)	Protein (aa)	Start codon	Stop codon	D1(bp)	D2(bp)	D3(bp)	D4(bp)
<i>psaA</i>	2	PS I	3,005	732	ATA*	TAA	73	47	112	120
<i>psaB</i>	3	PS I	3,121	776	ATA†	TAG	32	75	103	126
<i>psbA</i>	5	PS II	2,151	348	TTG	TAA	176	66	92	316
<i>psbB</i>	3	PS II	2,286	505	ATA	TAA	56	66	83	108
<i>psbC</i>	1	PS II	2,330	460	ATA*	TAG	100	36	118	239
<i>atpA</i>	2	ATPase	2,444	452	ATA	TAA	292	36	113	190
<i>petB</i>	1	Cyt b6f	2,204	219	ATG	TAA	818	65	90	117
23S rRNA‡	8		3,027				n.d.	86	99	n.d.
16S rRNA	1		2,563				n.d.	66	106	n.d.

n.d., Not determinable; PS, photosystem.

* Or TTG.

† Or ATT.

‡ There appear to be three related circles: two have deletions in the tripartite region, making them 25 bp and 35 bp shorter than the third.

(pUC18) libraries were made from DNA fragments of >0.5 kilobases (kb), 1.6–2 kb and 2–5 kb. Sequencing of 127 randomly picked clones from both ends yielded 48 clones (37%) containing homologues to known chloroplast genes, including large- and small-subunit chloroplast ribosomal RNA and seven chloroplast protein genes (Table 1). The other clones gave no significant hits and did not form contigs with the chloroplast genes. After contig assembly, chloroplast gene sequences were completed and confirmed using polymerase chain reaction (PCR) products amplified from genomic DNA.

The seven protein-coding genes are chloroplast-encoded in all known chloroplast genomes^{7,8}; they include at least one polypeptide from each of the four major membrane-protein assemblies of thylakoids (Table 1). We also found three circular contigs with truncated genes: two *psbC* genes and one 16S rRNA gene. Phylogenetic analyses of 23S rRNA sequences from chloroplasts, mitochondria and bacteria by several methods (neighbour joining, maximum parsimony and maximum likelihood) show that the *H. triquetra* 23S rRNA sequence lies solidly within the chloroplast lineage (Z.Z. et al., manuscript in preparation), eliminating the possibility that it is of bacterial or mitochondrial rather than plastid origin. We found no genes for transfer RNAs, ribosomal proteins or RNA polymerase. Sequencing extra clones from the *Sau3A* libraries gave more copies of the same nine genes. These results, and the earlier evidence for loss of plastid-encoded Rubisco^{9,10}, indicate that the chloroplast genome of *H. triquetra* may have an unusually reduced gene complement. However, Southern blotting with a spinach *psbD* probe revealed the presence of this gene in the *H. triquetra* satellite DNA band (Z.Z., unpublished) and in a small molecule fraction from *Amphidinium carterae* (R. G. Hiller, unpublished). A more exhaustive search for additional plastid genes is required to determine the chloroplast genome size.

Contig assembly showed that the sequence of each gene and its adjacent non-coding region is circularly permuted on different clones. To exclude the possibility of chimaeric cloning artefacts, we used inwardly and outwardly directed PCR primer pairs (Fig. 2)

to amplify DNA from uncloned genomic DNA. Both bA7/bA6 and bA1/bA5 primer pairs gave products of the size predicted for a circular contig. For each of the nine genes, sequences from both PCR products and clones could be integrated into a single circular contig, as shown for *psbA* in Fig. 2. We never found two different genes on the same clone or contig, even those that are adjacent in all other chloroplast genomes (for example, *psaA* and *psaB*). PCR using one primer from *psaA* and one from *psaB* gave no products; a negative result was also obtained with primers for 16S and 23S rRNA genes, which are adjacent and co-transcribed in all other plastids.

In principle, a circular contig could result either from a genuine physical circle containing only the gene and its adjacent non-coding region, or from identical tandem repeats of the gene with an intervening non-coding region. Hybridization of a spinach *psbA* probe with *EcoRI*-cut satellite DNA from *H. triquetra* labelled only a band at 2.1 kb (Fig. 1, lane 2), in agreement with the 2,151-base-pair (bp) size and single *EcoRI* site found by sequencing. If this gene existed as a tandem repeat within a larger molecule, digestion should have given fragments of at least two different sizes unless the number of copies was very large. When total *H. triquetra* DNA (uncut) was hybridized with a 0.7-kb probe from the *H. triquetra* 23S rRNA gene, it gave strong bands at about 2.1 and 4.3 kb, and a weaker band at 3.4 kb (Fig. 1, lane 14). The 2.1- and 4.3-kb bands disappeared after *BglII* digestion, giving two new bands of 0.6 and 2.6 kb (Fig. 1, lane 15), which sum to 3.2 kb, in reasonable agreement with the 3,027 bp predicted for a linear monomer from sequencing (Table 1). (The linear band is slightly smaller after digestion because of two closely spaced *BglII* sites.) The three bands in uncut DNA therefore probably correspond (in decreasing order of size) to relaxed monomeric circles, linear monomers and supercoiled circles. The weak hybridization at high molecular weight (>12 kb) is probably nonspecific crossreaction with nuclear 28S rRNA genes, because it does not decrease after digestion with *BglII*.

The non-coding region of each gene circle has a characteristic tripartite conserved region that we call 9G-9A-9G (Fig. 3). The

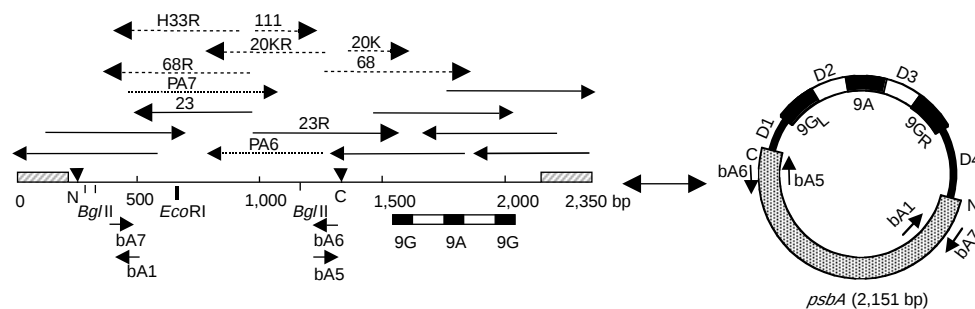


Figure 2 Structure of *psbA* contig assembled from 17 sequences (circularized on the right). Solid lines, sequences from clone 23 using M13 and walking primers; bold broken lines, end sequences of clones 68, 20K, 111 and 33; PA6, PA7, sequences from PCR products using bA6/bA7 primers. N, C, putative N and carboxy termini; grey rectangles, overlap regions of the linearized contig. *EcoRI* and *BglII* sites are indicated.

sequences from PCR products using bA6/bA7 primers. N, C, putative N and carboxy termini; grey rectangles, overlap regions of the linearized contig. *EcoRI* and *BglII* sites are indicated.

AAATCCTgAT AAATTTCACT TTTCTCAGTA cTTTTCCCCG GTAAAA**GGGG GGGGG**TGTCT
GCGATTTCAA AGTGGAGTCC CAAACGCATg TctGGAATAT ATGAGGagAa GTTATTTTCT
CAGATATTct CAGA---aAA AGTGCATTTT TGCACGATTT TaagAAAACA CATGCAATTT
GCCTTgAtt ttggccAagC CGCAAGTCCT TTA**AAAAAAA**ACaCTCGTG AAtGtTtttC
tATCATCTA TcATACCACC ttTGGTGGA TtATAGATAG ATTgctAgga TT-CACGAGT
TGAATAAAAA AGGGGAAaAt aATTt---a AATCCTGATA AATTTGACTT TTTCTCAGTAC
TTTTCCCCG TAAA**GGGG GGGG**TGTCTG CGATTTCAA GTGGAGTCCC AAACGCATgT
ctGGAATATA TGAGGaggag TAAAAATtTg ggAAgtttTa gaaa

Figure 3 Consensus sequence of the 9G-9A-9G region. Upper-case letters are nucleotides that are identical in all nine genes; lower case letters are nucleotides that are conserved in >50% of sequences. 9G and 9A motifs are boxed. Three dashes show positions of the D2 and D3 regions in Fig. 2.

central core of this region comprises 188 highly conserved nucleotides centred on a motif of nine adenines. It is flanked by two non-identical but closely related tracts of 135 highly conserved nucleotides, each centred on a run of nine guanines, and separated from the central 188-bp 9A tract by short, less well conserved regions (D2 and D3). The left (9G_L) and the right (9G_R) 135-bp 9G core regions have 107 nucleotides out of 135 in common. The 9G-9A-9G region of each of the nine genes is unique but strongly related to the others. The amino terminus of all the protein genes and the 5' ends of the sense strands of both rRNA genes have the same orientation with respect to the 9G-9A-9G region.

Comparison of all nine *H. triquetra* 9G-9A-9G regions shows that the 9G_L, 9A and 9G_R core regions are strongly conserved, although some substitutions and single-base insertions have occurred (Fig. 3). In contrast, the D2 region is more variable, with different numbers of a decanucleotide repeat (Fig. 4). In the 23S rRNA circle, there are six decanucleotide repeats, followed by a seventh with a two-nucleotide substitution. *petB* and *psaB* share an identical 43-nucleotide sequence (boxed in Fig. 4), and three genes share a 14-nucleotide variant in the D3 region (data not shown). Such variants being shared by two or more genes is indicative of gene conversion, which is probably the main homogenizing force between the 9G-9A-9G regions.

The entire conserved core of each non-coding region of *H. triquetra* circles can be folded into an elaborate secondary structure using RNA mfold (at <http://mfold.wustl.edu/~folder/rna>), but many equally stable structures are possible. As short repeats and A+T-rich sequences with a high potential for hairpins also occur in the replication origins of chloroplast DNA in *Euglena*¹¹ and *Chlamydomonas*¹², the tripartite regions might be the replication origins of the unigenic circles. These regions may also serve to bind the circles to a larger structure, such as a membrane, to mediate DNA segregation in order to avoid gene loss during chloroplast division.

RNA blots show that the chloroplast genes are expressed in messenger RNAs of the size expected for monocistronic transcripts (Fig. 5). The largest band is *psaB* (about 2.4 kb), which is only slightly larger than the minimum expected for a protein of 776 amino acids. There is no transcript large enough to encode both *psaA* and *psaB*, supporting the idea that these transcripts originate from single-gene circles and not from a larger DNA molecule carrying the typical bicistronic operon. Similar conclusions can be drawn for the *petB*, *psaA* and 23S rRNA transcripts. This does not

rule out the possibility that a larger circle carrying more than one gene exists, since monocistronic messages could still be transcribed from such a circle.

Five protein genes in *H. triquetra* use TAA as a stop codon and the other two use TAG; no TGA codons are used either as stops or as alternative amino-acid codons. Determining the start site was more difficult, as only *petB* has ATG at the position expected from alignment of the deduced protein sequence with plastid and cyanobacterial homologues. The *psbA* start codon is probably TTG as in *Chlorella* plastid *infA* (ref. 13), as the *H. pygmaea* and *H. rotundata* genes have ATG at the identical position (Z.Z. et al., manuscript in preparation). *psaA* has a TTG at a putative start site, but also has a nearby ATA, as used in some mitochondria¹⁴. *psbB* has ATA at the exact presumptive start position; *psaB* and *atpA* have an ATA a short distance upstream or downstream of their putative start sites. Moreover, even the *petB* gene has an ATA just upstream of its initial ATG. Although it is unclear whether chloroplast translation initiation requires Shine-Dalgarno sequences¹⁵, a GGTGG consensus motif, complementary to CCACC at the 3' end of the 16S rRNA sequence, is found at variable distances (-25 to -188 bp) upstream of the putative start codons of several genes.

Except at the N terminus, amino-acid sequences deduced from the seven protein-coding genes are very well aligned with the corresponding proteins from other organisms, but they are by far the most divergent. We have carried out phylogenetic analyses on the seven protein sequences concatenated to give an alignment of 3,302 amino acids and compared to concatenated sequences from plastids of thirteen other species⁷ (Fig. 6). *H. triquetra* groups with the diatom (*Odontella*), cryptomonad (*Guillardia*) and red alga (*Porphyra*), with 86% bootstrap support in neighbour joining and 51% in the parsimony tree. The tree shows that dinoflagellate plastids belong to the red algal/chromist lineage, although the branching order within this lineage is unclear. These data, along with the pigment composition and three-membrane plastid envelope, support the conclusion that typical dinoflagellate chloroplasts arose by secondary rather than primary symbiogenesis⁴⁻⁶.

To test the generality of this unusual chloroplast gene organisation, spinach *psbA* and *H. triquetra* 23S rRNA probes were hybridized to Southern blots of uncut total DNA from axenic *A. carterae* and three non-axenic dinoflagellates (Fig. 1). In *A. carterae*, *H. pygmaea* and *H. rotundata*, *psbA* exclusively labelled the 1.5–3 kb region (Fig. 1, lanes 6–9), consistent with circular DNA molecules in the same size range as in *H. triquetra*. Note that there is no label at the expected position of linearized minicircles (compare lane 2 with lanes 6–9), indicating that the upper and lower bands may be relaxed and supercoiled minicircles, respectively. The only significant labelling of higher molecular weight DNA was seen in *Gonyaulax* (Fig. 1, lane 5), which showed faint bands between 1.5 and 4 kb, and a slightly stronger band at 5.8 kb; whether the heavy labelling in the region above 12 kb is due to trapping of minicircles in the bulk DNA, or indicates the presence of *psbA* genes on a larger chromosome, is being investigated. Hybridization of the *H. triquetra* 23S rRNA gene probe to uncut DNA (Fig. 1, lanes 10–14) shows three bands in the 2–6 kb size range for all species except *Amphidinium*. To verify that *H. pygmaea* 23S rRNA is also on minicircles, its DNA was PCR-amplified using inwardly and

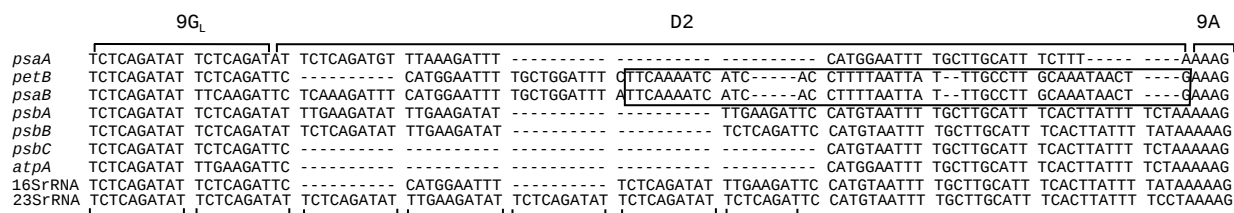


Figure 4 Alignment of the variable D2 region. Direct repeats (square brackets) and the gene conversion shared by *petB* and *psaB* (boxed) are shown.

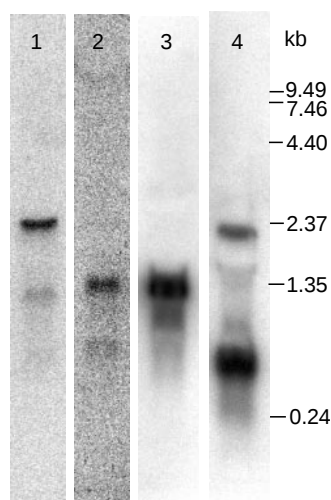


Figure 5 Hybridization of *H. triquetra* RNA with chloroplast gene probes. Lane 1, spinach *psaB*; lane 2, spinach *petB*; lane 3, *H. triquetra* *psbA*; lane 4, *H. triquetra* 23S rRNA.

outwardly directed primer pairs based on the *H. triquetra* sequence. The gene sequence is almost identical to those of the other two *Heterocapsa* species, but the tripartite non-coding region cannot be aligned with those of *H. triquetra* and *H. rotundata*. These results are consistent with the interpretation that, in all five species, most of the 23S rRNA genes are organized as monomeric circles.

An earlier study³ of *H. pygmaea* (identified there as *Glenodinium* sp.) estimated a chloroplast genome size of about 120 kb, based on restriction digests of satellite DNA, and reported *psbA*-labelled bands of 2.4 and 1.9 kb. We were unable to confirm this, but uncut *H. pygmaea* DNA shows four *psbA* bands of 3.0, 2.7, 1.4 and 1.2 kb (Fig. 1, lane 8), indicating that there may be two dissimilarly sized *psbA* minicircles. PCR amplification from *H. pygmaea* DNA using the ba6/ba7 primer pair (Fig. 2) yielded a single 0.8-kb band, but outward-directed primer pair ba1/ba5 yielded two bands of 1.7 and 1.9 kb. This shows that there are two circular *psbA* genes in *H. pygmaea*, and that the size heterogeneity is in the non-coding region; this was confirmed by sequencing. Similar PCR amplifications from *H. rotundata* gave one product each, and completion of the sequence gave a circular *psbA* contig of 2,298 bp. Our results show that unigenic circles are not restricted to *H. triquetra*. As *Amphidinium* and *Gonyaulax* are maximally divergent from *Heterocapsa* on the nuclear 18S rRNA tree¹⁶, this implies that this novel genome organization had already arisen in the common ancestor of all peridinin-containing dinoflagellate chloroplasts.

Earlier reports of subgenomic circles from chloroplasts concerned plasmid-like molecules that were present together with the main plastid chromosome and contained undefined sequences¹⁷ or only gene fragments¹⁸. The dinoflagellate organization is altogether different from these, as the minicircles contain only one gene. We cannot yet rule out the existence of a few chloroplast DNA molecules with a conventional multigenic organization, but the maintenance of two such differently organized genomes within the same organelle is evolutionarily unlikely, and would probably have generated populations of larger heterogeneous circles like those in the mitochondrial DNA of higher plants¹⁹. The unique organization of dinoflagellate plastid genes offers opportunities for studying the function and co-evolution of multiple replication origins in a single genome, and raises many issues for future study, such as the mechanism of copy-number control for the different circles and the evolution of minicircles.

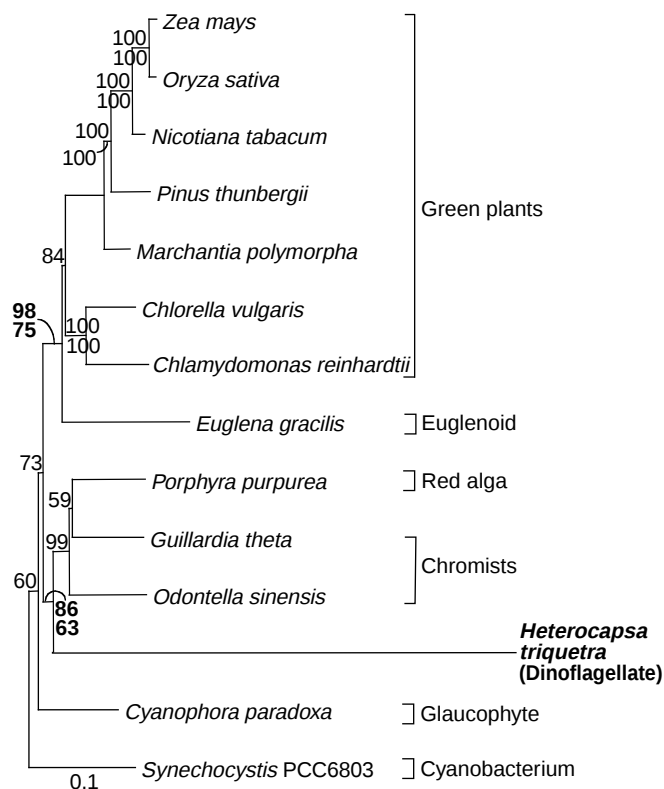


Figure 6 Neighbour-joining tree of seven concatenated protein sequences. Numbers on the branches are bootstrap values greater than 50% from neighbour joining (above) and maximum parsimony analyses (below). Scale represents 0.1 substitution per amino acid.

The dinoflagellate chloroplast DNA organization differs radically from all other plastid genomes^{6–8,13}, including the relict plastid of sporozoans²⁰ (such as the malarial parasite *Plasmodium falciparum*). Dinoflagellates and sporozoans are considered to be sister groups on the basis of nuclear rRNA sequences^{21,22}. If their plastids evolved by direct descent and divergence from a common ancestor (with which our unpublished 16S and 23S rRNA trees are consistent), it is likely that they share common protein-targeting mechanisms⁵ and perhaps other enzyme systems²³. This possibility needs to be investigated intensively; if it were confirmed, a comparison between the two kinds of alveolate plastid could have important implications for the development of antiplastid drugs to control malaria, toxoplasmosis and other important human and animal diseases caused by the sporozoa. □

Methods

Strains. *H. triquetra* (CCMP 449); *H. pygmaea* (CCMP 1490); *H. rotundata* (NEPCC D680); *A. carterae* (CCMP 1314); *Gonyaulax grindleyi* (NEPCC D535).

DNA and RNA isolation. Cells of axenic *H. triquetra* were suspended in lysis buffer (50 mM Tris-HCl, 100 mM EDTA, 100 mM NaCl, pH 8.0) at 4 °C, and broken by vortexing with glass beads. Incubation for 1 h at 50 °C with SDS (2%) and proteinase K (300 µg ml⁻¹) was followed by three phenol/chloroform extractions. For DNA³, Hoechst 33258 and CsCl were added and the solution centrifuged in a VTi 80 (Beckman) rotor at 55,000 r.p.m. for 24 h. Satellite and major DNA bands were recovered separately and recentrifuged. DNA was ethanol precipitated after removal of Hoechst 33258 by isopropanol extraction²⁴. Total DNA was isolated from the other species without CsCl purification. For RNA preparation, ethanol precipitation followed the phenol/CCl₄ extraction; RNA was selectively precipitated with LiCl²⁴.

Hybridization. DNA was electrophoresed on 0.9% agarose gels containing

Chlorophyll *b* and phycobilins in the common ancestor of cyanobacteria and chloroplasts

Akiko Tomitani^{*,†}, Kiyotaka Okada[‡], Hideaki Miyashita[§], Hans C. P. Matthijs^{||}, Terufumi Ohno[¶] & Ayumi Tanaka[†]

^{*} Department of Geology and Mineralogy, Graduate School of Science, Kyoto University, Kyoto 6068502, Japan

[†] Institute of Low Temperature Science, Hokkaido University, Sapporo 0600819, Japan

[‡] Department of Botany, Graduate School of Science, Kyoto University, Kyoto 6068502, Japan

[§] Marine Biotechnology Institute, Kamaishi Laboratories, Kamaishi, Iwate 0260001, Japan

^{||} Department of Microbiology, University of Amsterdam,

Nieuwe Achtergracht 127, NL 1018 WS, Amsterdam, The Netherlands

[¶] The Kyoto University Museum, Kyoto University, Kyoto 6068502, Japan

Photosynthetic organisms have a variety of accessory pigments, on which their classification has been based. Despite this variation, it is generally accepted that all chloroplasts are derived from a single cyanobacterial ancestor^{1–3}. How the pigment diversity has arisen is the key to revealing their evolutionary history. Prochlorophytes are prokaryotes which perform oxygenic photosynthesis using chlorophyll *b*, like land plants and green algae (Chlorophyta), and were proposed to be the ancestors of chlorophyte chloroplasts^{4,5}. However, three known prochlorophytes (*Prochloron didemni*, *Prochlorothrix hollandica* and *Prochlorococcus marinus*) have been shown to be not the specific ancestors of chloroplasts, but only diverged members of the cyanobacteria, which contain phycobilins but lack chlorophyll *b*^{6,7}. Consequently it has been proposed that the ability to synthesize chlorophyll *b* developed independently several times in prochlorophytes and in the ancestor of chlorophytes. Here we have isolated the chlorophyll *b* synthesis genes (chlorophyll *a* oxygenase)⁸ from two prochlorophytes and from major groups of chlorophytes. Phylogenetic analyses show that these genes share a common evolutionary origin. This indicates that the progenitors of oxygenic photosynthetic bacteria, including the ancestor of chloroplasts, had both chlorophyll *b* and phycobilins.

The chlorophyll *a* oxygenase (CAO) gene is involved in chlorophyll *b* biosynthesis, and was isolated by insertional mutagenesis of the green alga *Chlamydomonas reinhardtii*⁸. This gene encodes a protein of 463 amino acids, including a putative transit peptide. The deduced amino-acid sequence contains binding domains for a [2Fe-2S] Rieske centre and for a mononuclear non-haem iron, indicating that CAO encodes an oxygenase. When a CAO complementary DNA was isolated from *Arabidopsis thaliana* and expressed in *Escherichia coli*, enzymatic study with the expressed protein showed that conversion of chlorophyll *a* to chlorophyll *b* was carried out by the CAO product alone⁹. All oxygenic photosynthetic organisms contain chlorophyll *a* and can synthesize chlorophyll *b* only by obtaining CAO. Analyses of chlorophyll *b* less mutants of *Chlamydomonas* indicate that CAO is a single-copy gene⁸. Comparison of the deduced amino-acid sequences of the *Chlamydomonas* and *Arabidopsis* CAOs showed that a region of about 300 amino-acid residues, which includes two motifs for a [2Fe-2S] Rieske centre and for a mononuclear iron, is highly conserved. On the other hand, both the amino- and carboxy-terminal regions have low similarity.

We isolated CAO genes or cDNAs from *P. hollandica*, *P. didemni* (Prochlorophyta), *Dunaliella salina* (Chlorophyceae), *Marchantia polymorpha* (Bryophyta), and *Oryza sativa* (Angiospermae) by

200 ng ml⁻¹ ethidium bromide; RNA was electrophoresed on 1.2% agarose-formaldehyde gels²⁴. Probes were labelled by a random primer system (Life Technologies) and [³²P]dCTP (Amersham). Membranes were hybridized at 55 °C in Church buffer (0.25 M Na₂HPO₄, 1 mM EDTA, 7% SDS), washed with 0.1 × SSC/0.1% SDS at 55 °C (DNA) or 65 °C (RNA), and exposed to Kodak X-Omat film.

DNA sequencing and analysis. Sequencing was done using a dye terminator cycle sequencing kit (Applied Biosystems); traces from the automatic sequencer were edited by Staden's Trev program (<http://www.mrc-lmb.cam.ac.uk/pubseq>). PCR reactions (Sigma reagents) were carried out for 35 cycles: 94 °C for 30 s, 55 °C for 30 s, followed by 2 min at 72 °C. PCR products purified from 1% low-melting-agarose gels were used in sequencing. A database was set up in GAP4 of Staden and contigs were generated. We retrieved protein sequence alignments⁷ from 134.169.70.80/ftp/pub/incoming/; and *Guillardia* sequences⁸ from Genbank (AF041468). Alignments were improved manually in GDE (ver. 2.2)²⁵. PHYLIP (ver. 3.5) (<http://evolution.genetics.washington.edu/phylip/html>) was used to construct a maximum parsimony tree with global rearrangement, and neighbour-joining trees using the PAM matrix. The input order of taxa was jumbled.

Received 25 January; accepted 29 April 1999.

1. Dodge, J. D. A survey of chloroplast ultrastructure in Dinophyceae. *Phycologia* **14**, 253–263 (1975).
2. Jeffrey, S. W. et al. Chloroplast pigment patterns in dinoflagellates. *J. Phycol.* **111**, 374–384 (1975).
3. Boczar, B. A., Liston, J. & Cattolico, R. A. Characterization of satellite DNA from three marine dinoflagellates: *Glenodinium* sp. and two members of the toxic genus, *Protogonyaulax*. *Plant Physiol.* **97**, 613–618 (1991).
4. Gibbs, S. P. The chloroplasts of some groups of algae may have evolved from endosymbiotic eukaryotic algae. *Ann. NY Acad. Sci.* **361**, 193–207 (1981).
5. Cavalier-Smith, T. in *Biodiversity and Evolution* (eds Arai, R., Kato, M. & Doi, Y.) 75–114 (The National Science Museum Foundation, Tokyo, 1995).
6. Palmer, J. D. & Delwiche, C. F. in *Molecular Systematics of Plants II* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 375–409 (Kluwer, Norwell, MA, 1998).
7. Martin, W. et al. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165 (1998).
8. Douglas, S. E. & Penny, S. L. The plastid genome from the cryptomonad alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J. Mol. Evol.* **48**, 236–244 (1999).
9. Morse, D., Salois, P., Markovic, P. & Hastings, J. W. A nuclear encoded form II RuBisCo in dinoflagellates. *Science* **268**, 1622–1624 (1995).
10. Rowan, R., Whitney, S. W., Fowler, A. & Yellowless, D. Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs, encoded by a nuclear multi-gene family. *Plant Cell* **8**, 539–553 (1996).
11. Schlunegger, B. & Stutz, E. The *Euglena gracilis* chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication. *Curr. Genet.* **8**, 629–634 (1984).
12. Wu, M., Lou, J. K., Chang, D. Y., Chang, C. H. & Nie, Z. Q. Structure and function of a chloroplast DNA replication origin of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA* **83**, 6761–6765 (1986).
13. Wakasugi, T. et al. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc. Natl Acad. Sci. USA* **94**, 5967–5972 (1997).
14. Boore, J. L. & Brown, W. M. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. *Genetics* **138**, 423–443 (1994).
15. Sugiura, M., Hirose, T. & Sugita, M. Evolution and mechanism of translation in chloroplasts. *Annu. Rev. Genet.* **32**, 437–459 (1998).
16. Saunders, G. W., Hill, D. R. A., Sexton, J. P. & Andersen, R. A. Small-subunit ribosomal RNA sequences from selected dinoflagellates: testing classical evolutionary hypotheses with molecular systematic methods. *Plant Syst. Evol.* (suppl.) **11**, 237–259 (1997).
17. Jacobs, J. D. et al. Characterisation of two circular plasmids from the marine diatom *Cylindrotheca fusiformis*: plasmids hybridise to chloroplast and nuclear DNA. *Mol. Gen. Genet.* **233**, 302–310 (1992).
18. La Claire II, J. W., Loudenslager, C. M. & Zuccarello, G. C. Characterization of novel extrachromosomal DNA from giant celled marine green algae. *Curr. Genet.* **34**, 204–211 (1998).
19. Backert, S., Nielsen, B. L. & Börner, T. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci.* **2**, 477–483 (1987).
20. Wilson, R. J. M. et al. Complete map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155–172 (1996).
21. Gajadhar, A. A. et al. Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates and ciliates. *Mol. Biochem. Parasitol.* **45**, 147–154 (1991).
22. Cavalier-Smith, T. Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994 (1993).
23. Waller, R. et al. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **98**, 12352–12357 (1998).
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning, A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).
25. Smith, S. W. The genetic data environment and expandable GUI for multiple sequence analysis. *Comp. Appl. Biosci.* **10**, 671–675 (1994).

Acknowledgements. We thank M. Beaton and K. Ishida for valuable discussion and advice; E. Filek for help with plasmid sequencing; X. Wu and E. Chao for advice on PCR; R. G. Hiller for communicating unpublished data; and J. Saldarriaga for *H. rotundata* and *G. grindleyi* total DNA. This work was supported by NSERC research grants to B.R.G. and T.C.-S. T.C.-S. thanks the Canadian Institute for Advanced Research for fellowship support.

Correspondence and requests for materials should be addressed to B.R.G. (e-mail: brgreen@interchange.ubc.ca). The sequences reported here will appear in the Genbank sequence database under accession numbers AF 130031–AF 130039.

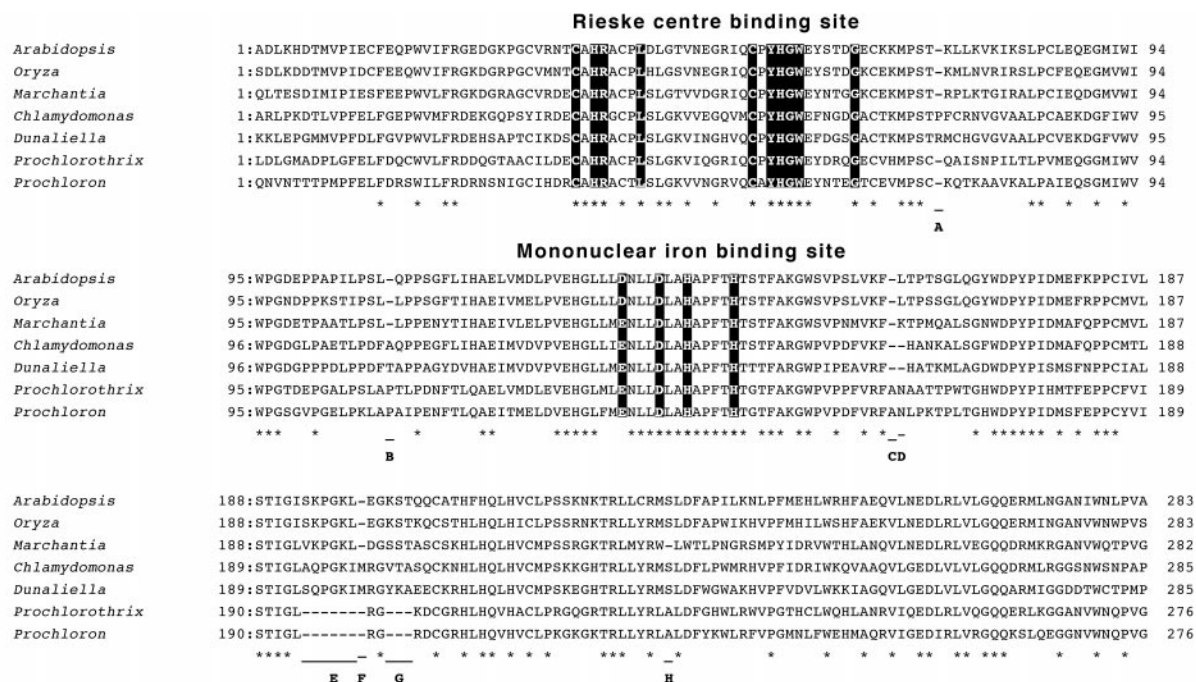


Figure 1 Alignment of CAO amino-acid sequences from chlorophytes *A. thaliana*, *O. sativa*, *M. polymorpha*, *C. reinhardtii* and *D. salina*, and prochlorophytes *P. hollandica* and *P. didemni*. Asterisks indicate conserved identical amino-acid

residues. The residues of the two binding sites for a [2Fe-2S] Rieske centre and for a mononuclear iron are highlighted in black. The eight sites where amino-acid gaps exist in the alignment are underlined and labelled A to H.

polymerase chain reaction (PCR) using DNA from genomes or cDNAs as templates or by screening cDNA libraries. We then determined the nucleotide sequences of the highly conserved region. The deduced amino-acid sequences consisted of almost the same number of residues (276–285 amino acids). Their alignment is presented in Fig. 1. All sequences shared 106 identical amino-acid residues (indicated by asterisks in Fig. 1) and had the two motifs for a [2Fe-2S] Rieske centre and for a mononuclear iron (highlighted in black in Fig. 1). Any two of the CAO sequences showed 51 to 83% amino-acid identity.

There were eight sites where amino-acid gaps existed in the alignment (A–H in Fig. 1). The members of each taxon (Prochlorophyta, Chlorophyceae, Bryophyta and Angiospermae) shared the same site(s) of these gaps, indicating that these sites could be used as phylogenetic markers. Gap sites A, B, C, and F were shared: A and F were shared by Angiospermae, Bryophyta and Prochlorophyta; B by Angiospermae and Bryophyta; and C by Angiospermae, Bryophyta and Chlorophyceae. The other gap sites, D, E, G and H, were each characteristic of one taxon: D was characteristic of Chlorophyceae, E and G of Prochlorophyta, and H of Bryophyta.

To determine whether the CAO genes shared a common origin and which sequence should be used as an outgroup for detailed analysis of CAO, we retrieved about 100 protein sequences which showed similarity to CAO sequences, and constructed a phylogenetic tree by the distance matrix method^{10,11} (data not shown). The tree showed that the CAO sequences formed a cluster, in which the other sequences were not included. In this tree, the closest known relative of the CAO sequences was Tic55 from *Pisum sativum*¹², which made a cluster with slr1747¹³ from *Synechocystis* sp. and with lsl1 (ref. 14) from higher plants.

To examine the relationship between prochlorophytes and chlorophytes, we constructed phylogenetic trees using the amino-acid sequences deduced from the CAO genes, with the Tic55 protein sequence as an outgroup, by the distance matrix method^{10,11} (Fig. 2) and the maximum likelihood method^{15,16} (data not shown). Each method gave the same tree topology. The members of each taxon formed a cluster and the branching pattern, which was similar to those inferred from *psbA*¹⁷, 16S ribosomal RNA¹⁸ and *rpoC1* (ref. 6)

sequences, reflected the evolutionary relationships between Prochlorophyta, Chlorophyceae, Bryophyta and Angiospermae (Fig. 2). The tree topology was supported by the resampling bootstrap method¹⁹ with probabilities of 93–100% by the distance matrix method^{10,11} and 66–100% by the maximum likelihood method^{15,16}. The *Prochlorothrix* and *Prochloron* CAO sequences shared 69% identical amino-acid residues and clustered with high bootstrap probabilities of 100% by the distance matrix method and 93% by the maximum likelihood method.

We have shown that the *Prochlorothrix* and *Prochloron* CAOs have high sequence similarity, including gap positions characteristic of Prochlorophyta (Fig. 1), and that *Prochlorothrix* and *Prochloron* are the nearest neighbours to each other in the phylogenetic tree (Fig. 2). These results strongly indicate that the ability to synthesize chlorophyll *b* did not arise independently several times in each of the prochlorophyte lineages as considered so far^{6,7}, but shares a common origin. This was also implied by analyses of prochlorophyte light-harvesting systems. Most eukaryotic light-harvesting

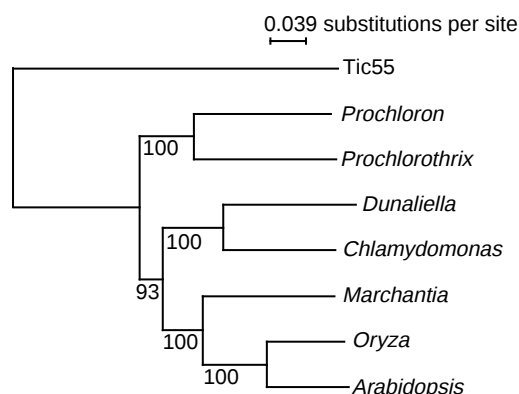


Figure 2 Phylogenetic relationships of prochlorophytes and chlorophytes inferred from CAO amino-acid sequences. The tree was constructed by the neighbour-joining method^{10,11}. The deepest root was determined using Tic55¹² as an outgroup. Numbers at the branch points represent the bootstrap values for percentage of 1,000 replicate trees.

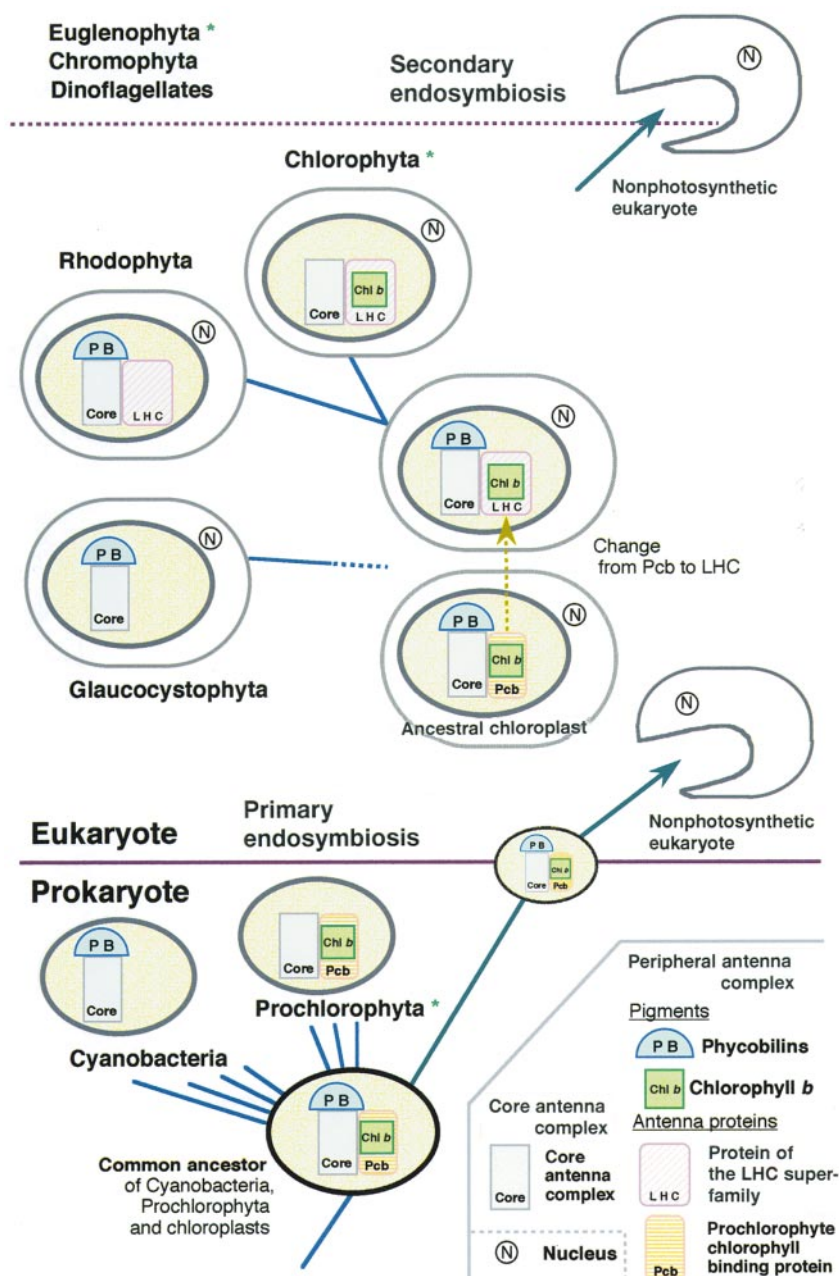


Figure 3 Hypothetical scheme for the evolution of oxygenic photosynthetic prokaryotes and eukaryotes from the common ancestor containing chlorophyll *b* and phycobilins. Colour marks designating core antenna complexes, peripheral antenna complexes (antenna proteins and pigments), and nucleus are given at lower right. Solid arrows indicate primary or secondary endosymbiosis. Asterisks (*) indicate chlorophyll-*b*-containing organisms. Cyanobacteria have no LHCs¹³ containing three membrane-spanning helices, but have HLIP containing a single

membrane-spanning α -helix related to LHCs²⁴. LHCs have been identified in Rhodophyta²⁵ but not in Glaucocystophyta. The ancestral LHCs are considered to have been present before the divergence into Rhodophyta and Chlorophyta²⁵. Only three algal groups are shown to represent the photosynthetic eukaryotes with secondary chloroplasts. Photosynthetic organisms represented in this scheme without taxon names are hypothetical ones.

chlorophyll protein complexes (LHCs) are members of an extended protein family (the LHC superfamily), including the chlorophyll *a/b*-binding proteins (CABs)²⁰. However, all of the three prochlorophytes use chlorophyll *a/b* proteins (Pcb), which do not belong to the LHC superfamily but are closely related to IsiA²¹. Assuming multiple evolutionary origins of prochlorophytes in cyanobacterial radiation^{6,7}, it can be considered that prochlorophytes and cyanobacteria derived from hypothetical ancestors of oxygenic photosynthetic bacteria containing both chlorophyll *b* and phycobilins (Fig. 3). Subsequent losses of chlorophyll *b* or phycobilins from the ancestors would have led to the cyanobacteria or prochlorophyte lineages, respectively (Fig. 3). *Prochlorococcus marinus* is the sole

photosynthetic organism known so far that contains (divinyl-) chlorophyll *b* and a type of phycoerythrin²². This organism might have derived from an ancestral form without major pigmentary alternations and could be a model organism for the ancestral oxygenic photosynthetic bacteria.

Furthermore, the CAOs from prochlorophytes and chlorophytes have high sequence similarity (Fig. 1), indicating that their chlorophyll *b* synthesis genes derived from a common evolutionary origin. This is confirmed by their phylogenetic relationships, represented in the tree (Fig. 2). All chloroplasts have been generally accepted to have a monophyletic origin¹⁻³, and phycobilins are the only identified accessory pigments that are shared by both photosynthetic

prokaryotes (cyanobacteria) and eukaryotes (Rhodophyta (red algae) and Glaucocystophyta)). From these observations it has been believed that the ancestral chloroplast was rhodophyte- (and glaucocystophyte-) like in containing phycobilins, as do cyanobacteria, and that subsequently chlorophytes acquired chlorophyll *b* and lost phycobilins. Taking into consideration our finding that chlorophyll *b* in prochlorophytes and chlorophytes has a common evolutionary origin, it is more reasonable to assume that the origin of chloroplasts were oxygenic photosynthetic bacteria containing chlorophyll *b* and phycobilins, which would have been derived from the hypothetical common ancestor of prochlorophytes and cyanobacteria proposed here (Fig. 3). Therefore, the ancestral photosynthetic eukaryotes should have possessed both chlorophyll *b* and phycobilins (Fig. 3). This is consistent with the fact that all the algal groups that are believed to have primary-endosymbiotic chloroplasts (Chlorophyta, Rhodophyta and Glaucocystophyta)^{1,2} contain either chlorophyll *b* or phycobilins as accessory pigments (Fig. 3). In the ancestral photosynthetic eukaryotes, chlorophyll *b* might have been bound to Pcb, and then transferred to CABs, which arose soon after the primary endosymbiosis (Fig. 3). Subsequently chlorophyll *b* would have been lost in the lineages of Rhodophyta and Glaucocystophyta, while phycobilins were also lost in the Chlorophyta (Fig. 3). □

Methods

Gene isolation and sequencing. We isolated a full-length CAO cDNA clone of *A. thaliana* by screening an *Arabidopsis* cDNA library with an EST clone obtained from the Arabidopsis Biological Resource Center as a probe. A CAO cDNA clone of *O. sativa* was obtained from the Rice Genome Research Program. A *M. polymorpha* cDNA was a gift from H. Fukuzawa (Kyoto University, Japan). We extracted genomic DNA of *P. hollandica* from cultured cells and purified it on a CsCl gradient²³. Genomic DNA of *P. didemni* was extracted²³ from cells that were collected from an ascidian *Lissoclinum patella* and were frozen. We carried out control PCR experiments with primers for eukaryotic and *Prochloron* rRNA and obtained amplified rRNA genes using primers for *Prochloron* but not for eukaryotes; accordingly, there was no contamination with chlorophyll *b*-containing eukaryotes in the *Prochloron* cells. Parts of CAO cDNAs from *M. polymorpha* and *D. salina* were amplified by PCR using cDNAs as templates with degenerate primers to the regions conserved in both *C. reinhardtii* and *A. thaliana* CAO sequences. We obtained a CAO cDNA clone of *D. salina* by screening a cDNA library using the PCR product for *Dunaliella salina* as a probe. Parts of the CAO genes from *P. hollandica* and *P. didemni* were amplified by PCR using genomic DNAs as templates with the degenerate primers. The PCR products were cloned into a pBluescript plasmid vector (Stratagene). The nucleotide sequences were determined using the Dye Terminator DNA sequencing kit (Applied Biosystems) by a DNA sequencer (model 310, Applied Biosystems).

Phylogenetic analyses. The deduced amino-acid sequences of CAO and Tic55 (accession no. AJ000520) were aligned using CLUSTAL W¹⁰ with manual refinement. Phylogenetic trees were generated using CLUSTAL W¹⁰ for the distance matrix method¹¹ and using MOLPHY¹⁵ for the maximum likelihood method¹⁶. All the amino-acid sites where gaps exist in the alignment were excluded from the calculation for the tree presented here. The same tree topology was obtained when those gap-located sites were included in the calculation (data not shown). We obtained the same tree topology when we used either slr1747 (ref. 13) or *lls1* (ref. 14) sequence as an outgroup (data not shown).

Received 22 January; accepted 13 May 1999.

- Delwiche, C. F. & Palmer, J. D. in *Origins of the Algae and Their Plastids* (ed. Bhattacharya, D.) 53–96 (Springer, Wien, Germany, 1997).
- Bhattacharya, D. & Medlin, L. Algal phylogeny and the origin of land plants. *Plant Physiol.* **116**, 9–15 (1998).
- Wolfe, G. R. *et al.* Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation. *Nature* **367**, 566–568 (1994).
- Lewin, R. A. & Withers, N. W. Extraordinary pigment composition of a prokaryotic alga. *Nature* **256**, 735–737 (1975).
- Lewin, R. A. Prochlorophyta as a proposed new division of algae. *Nature* **261**, 697–698 (1976).
- Palenik, B. & Haselkorn, R. Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. *Nature* **355**, 265–267 (1992).
- Urbach, E., Robertson, D. L. & Chisholm, S. W. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* **355**, 267–270 (1992).

- Tanaka, A. *et al.* Chlorophyll *a* oxygenase (CAO) is involved in chlorophyll *b* formation from chlorophyll *a*. *Proc. Natl Acad. Sci. USA* **95**, 12719–12723 (1998).
- Rüdiger, W. *et al.* in *The Chloroplast: From Molecular Biology to Biotechnology* (eds Argyroudi-Akoyunoglou, J. H. & Senger, H.) 185–190 (Kluwer, Netherlands, 1999).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- Caliebe, A. *et al.* The chloroplastic protein import machinery contains a Rieske-type iron-sulfur cluster and a mononuclear iron-binding protein. *EMBO J.* **16**, 7342–7350 (1997).
- Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
- Gray, J., Close, P. S., Briggs, S. P. & Johal, G. S. A novel suppressor of cell death in plants encoded by the *lls1* gene of maize. *Cell* **89**, 25–31 (1997).
- Adachi, J. & Hasegawa, M. *Computer Science Monographs*, No. 28. MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood (Institute of Statistical Mathematics, Tokyo, 1996).
- Kishono, H., Miyata, T. & Hasegawa, M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **30**, 151–160 (1990).
- Morden, C. W. & Golden, S. S. *PsbA* genes indicate common ancestry of prochlorophytes and chloroplasts. *Nature* **337**, 382–385 (1989).
- Bhattacharya, D. & Medlin, L. The phylogeny of plastids: A review based on comparison of small-subunit ribosomal RNA coding regions. *J. Phycol.* **31**, 489–498 (1995).
- Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
- Green, B. R. & Pichersky, E. Hypothesis for the evolution of three-helix Chl *a/b* and Chl *a/c* light-harvesting antenna proteins from two-helix and four-helix ancestors. *Photosyn. Res.* **39**, 149–162 (1994).
- La Roche, J. *et al.* Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. *Proc. Natl Acad. Sci. USA* **93**, 15244–15248 (1996).
- Hess, W. R. *et al.* Coexistence of phycoerythrin and a chlorophyll *a/b* antenna in a marine prokaryote. *Proc. Natl Acad. Sci. USA* **93**, 11126–11130 (1996).
- Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology* 2.4 (Wiley, New York, 1987).
- Dolganov, N. A. M., Bhaya, D. & Grossman, A. R. Cyanobacterial protein with similarity to the chlorophyll *a/b* binding proteins of higher plants: Evolution and regulation. *Proc. Natl Acad. Sci. USA* **92**, 636–640 (1995).
- Tan, S., Cunningham, F. X. Jr & Grant, E. *LhcaR1* of the red alga *Porphyridium cruentum* encodes a polypeptide of the LHCl complex with seven potential chlorophyll *a*-binding residues that are conserved in most LHCS. *Plant. Mol. Biol.* **33**, 157–167 (1997).

Acknowledgements. We thank N. Iwabe for helpful discussions about phylogenetic analyses and comments on the manuscript; Y. Koshino for the use of unpublished sequences of *Arabidopsis thaliana*; A. Melis, M. Mimuro, and K. Shimizu for comments on the manuscript; R. Tanaka for technical assistance and comments on the manuscript; H. Fukuzawa for a *Marchantia polymorpha* cDNA; and T. Maruyama for helpful discussions.

Correspondence and requests for materials should be addressed to A. Tomitani (e-mail: tomitani@terra.kuops.kyoto-u.ac.jp). Sequences have been deposited in GenBank under accession numbers AB021310–021314 and AB021316.

Improved auditory spatial tuning in blind humans

Brigitte Röder*, Wolfgang Teder-Sälejärvi†, Anette Sterr‡, Frank Rösler*, Steven A. Hillyard† & Helen J. Neville§

* Department of Psychology, Philipps-University Marburg, Gutenbergstrasse 18, 35037 Marburg, Germany

† Department of Neurosciences, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, USA

‡ Department of Psychology, University of Konstanz, 78457 Konstanz, Germany

§ Department of Psychology, University of Oregon, Eugene, Oregon 97403-1227, USA

Despite reports of improved auditory discrimination capabilities in blind humans^{1–3} and visually deprived animals⁴, there is no general agreement as to the nature or pervasiveness of such compensatory sensory enhancements⁵. Neuroimaging studies have pointed out differences in cerebral organization between blind and sighted humans^{6–12}, but the relationship between these altered cortical activation patterns and auditory sensory acuity remains unclear. Here we compare behavioural and electrophysiological indices of spatial tuning within central and peripheral auditory space in congenitally blind and normally sighted but blindfolded adults to test the hypothesis (raised by earlier studies of the effects of auditory deprivation on visual processing^{13,14}) that the effects of visual deprivation might be more pronounced for processing peripheral sounds. We find that blind participants displayed localization abilities that were superior to those of sighted controls, but only when attending to sounds in peripheral

auditory space. Electrophysiological recordings obtained at the same time revealed sharper tuning of early spatial attention mechanisms in the blind subjects. Differences in the scalp distribution of brain electrical activity between the two groups suggest a compensatory reorganization of brain areas in the blind that may contribute to the improved spatial resolution for peripheral sound sources.

The ability to focus attention selectively on relevant external sound sources is essential for auditory perception. Behavioural studies have shown that auditory attention, like visual attention¹⁵, is located in space in the form of a gradient, with the most efficient sound processing occurring at the attended locations and a progressive decline at increasingly distant locations¹⁶. The brain mechanisms of auditory spatial attention have been studied extensively in humans by means of non-invasive scalp recordings of event-related potentials (ERPs)^{17,18}. In particular, focusing attention on a sound source in the environment is indexed by a negative ERP

beginning 80–100 ms after the onset of the stimulus (the N1 component), which is greater in amplitude for sounds at attended than at unattended locations. A recent study using both behavioural and ERP measures found that the spatial gradient of auditory attention was less narrowly focused when subjects attended to peripheral as opposed to central sound locations¹⁹, which is in agreement with previous behavioural studies showing decreases in auditory localization accuracy with increasing eccentricity of the sound source^{20,21}. We have used similar behavioural and ERP methods to test whether mechanisms of auditory spatial attention display compensatory enhancements in blind individuals, and whether these effects are more pronounced for peripheral than for central auditory space.

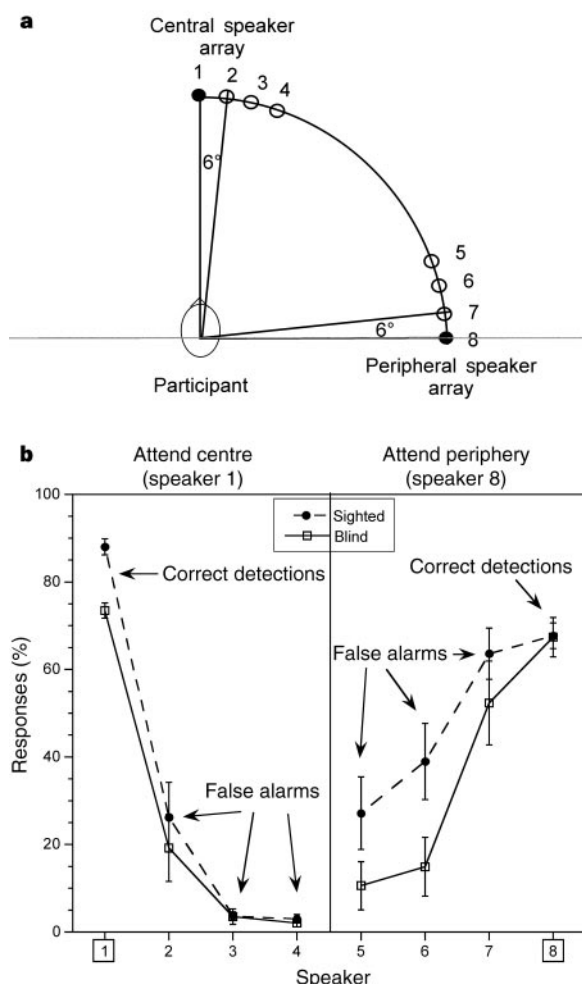


Figure 1 Speaker layout and response gradients. **a**, Central and peripheral speaker arrays. Participants had to detect rare deviants at speaker 1 (attend centre) or speaker 8 (attend periphery). **b**, Gradients of the percentage of detection responses (mean $\pm$ standard error) to deviants at the central speakers 1–4 and peripheral speakers 8–5 when the participant's task was to press a button to deviants at speaker 1 (attend centre) and speaker 8 (attend periphery), respectively. Responses to deviants at locations 1 and 8 were classified as correct responses, whereas responses to the remaining locations were considered false alarms. Response rates to deviants in the unattended speaker array were negligible and are not shown. Sighted and blind participants did not differ in their gradients of detection performance in the 'attend centre' condition, but blind participants showed a more sharply tuned gradient of attention than sighted subjects in the 'attend periphery' condition.

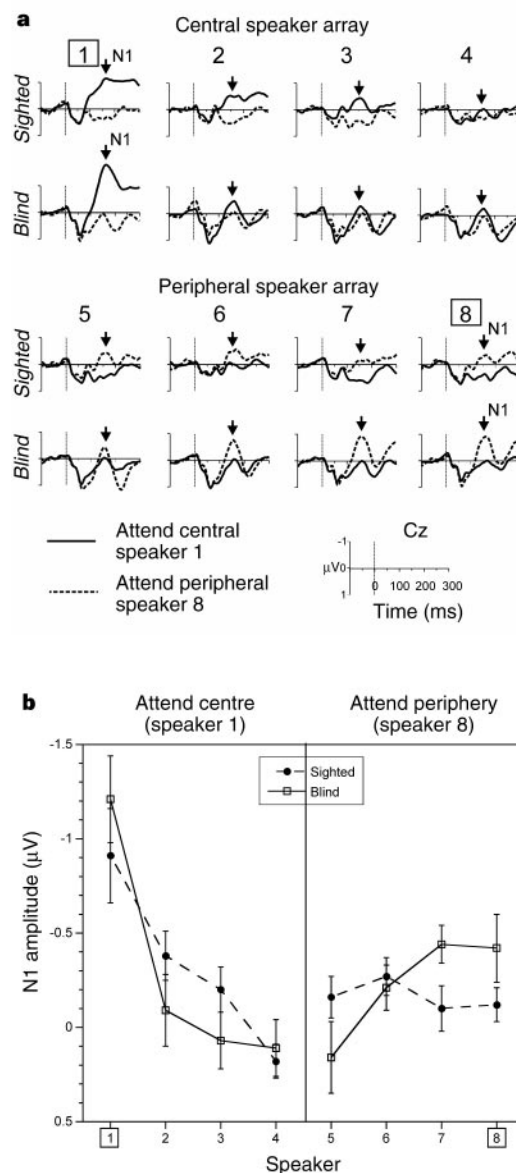


Figure 2 Event-related potentials to standard stimuli. **a**, ERPs recorded from the vertex (Cz) in response to standard stimuli when speaker 1 (solid line) and speaker 8 (dashed line) were attended. The N1 component is indicated by arrows above the waveforms. Because no behavioural responses were required to standards, the N1 amplitude reflects auditory spatial tuning without any contamination by motor activity. **b**, Gradients of mean N1 amplitude ($\pm$ s.e.) to standard stimuli at speakers within the attended central and the attended peripheral arrays. N1 was measured from 100 to 200 ms poststimulus. The early attention mechanism indexed by N1 was more sharply tuned in the blind than in the sighted subjects when they were attending to targets at the peripheral speaker 8.

Table 1 Subjects' characteristics

Participant	Age (years)	Sex	Age and cause of blindness	Visual perception	Handedness
1	26	Female	Birth, glaucoma	Diffuse light	Right
2	43	Male	Birth, ROP	No	Right
3	36	Female	<1 year, retinoblastoma	No	Right
4	41	Female	Birth, ROP	No	Right
5	45	Female	Birth, ROP	No	Right
6	41	Female	Birth, ROP	No	Right
7	31	Male	Birth, retinoblastoma	No	Right
8	46	Female	Birth, congenital eye abnormalities, born with small eyes	No	Right

Blind subjects: mean age, 39 years (standard deviation, 7; range, 26–46 years). Control subjects: mean age, 41 years (s.d. 8; range, 28–52 years), 6 female and 2 male, 4 with normal and 4 with corrected-to-normal vision, all right handed. Sighted participants were blindfolded. All participants reported normal hearing, which was confirmed in a subsample by a hearing test.

Auditory spatial attention was assessed using brief noise bursts presented in random sequence with equal probability from four central and four peripheral speakers (see Fig. 1). In half the runs, the subject's task was to detect infrequent 'deviant' sounds (of a higher pitch) at the central speaker 1 (the 'attend centre' condition), and in the other half the subject had to detect deviants at the peripheral speaker 8 (the 'attend periphery' condition). Sounds from all other speakers were to be ignored. Behavioural measures of target detection accuracy and concurrently recorded ERPs were obtained from eight congenitally blind subjects and eight blindfolded, sighted control subjects matched on age, handedness and gender (Table 1). In each run, ERPs were recorded and averaged separately to the frequent ('standard') and rare deviant noise bursts at each speaker location.

As shown in Fig. 1, in the 'attend centre' condition, all subjects were highly accurate at detecting the deviant sounds at the designated location (speaker 1), and response rates declined progressively as a function of distance from the attended speaker. Although the correct detection rate at speaker 1 was higher in the sighted than the blind subjects ($t(1, 14) = 5.7$, $P < 0.001$), the two groups did not differ significantly in the sharpness of attentional tuning, as reflected in the proportional decline of false alarm responses

made to the adjacent speaker locations ($P > 0.2$ for all inter-group comparisons). In the 'attend periphery' condition, both groups were less accurate and made more false alarms to the adjacent speakers (analysis of variance (ANOVA) results: speaker, $F(3, 42) = 112.56$, $P < 0.001$; condition, $F(1, 14) = 31.59$, $P < 0.001$; speaker $\times$ condition, $F(3, 42) = 21.15$, $P < 0.001$). Most importantly, however, the proportional decline in response rate between the attended and adjacent speakers was greater for the blind than for the sighted subjects, indicating a more narrow focusing of attention on the peripheral target location in the blind listeners. This steeper gradient of responding for the blind subjects across the peripheral but not the central array was reflected in a significant speaker $\times$ condition $\times$ group interaction ($F(3, 42) = 3.45$, $P < 0.04$) and in a sharper decline in false alarm rates in the blind group going from speaker 8 to 6 ($t(14) = 1.90$, $P < 0.05$) and from speaker 7 to 6 ($t(14) = 1.78$, $P < 0.05$).

The effects of spatial attention on early auditory processing were indexed by the amplitude of the N1 component of the ERP elicited by sounds from each speaker. In both groups, the N1 amplitude (mean voltage over 100–200 ms) decreased progressively in response to sounds increasingly distant from the attended speaker and, like the behavioural data, this gradient was steeper

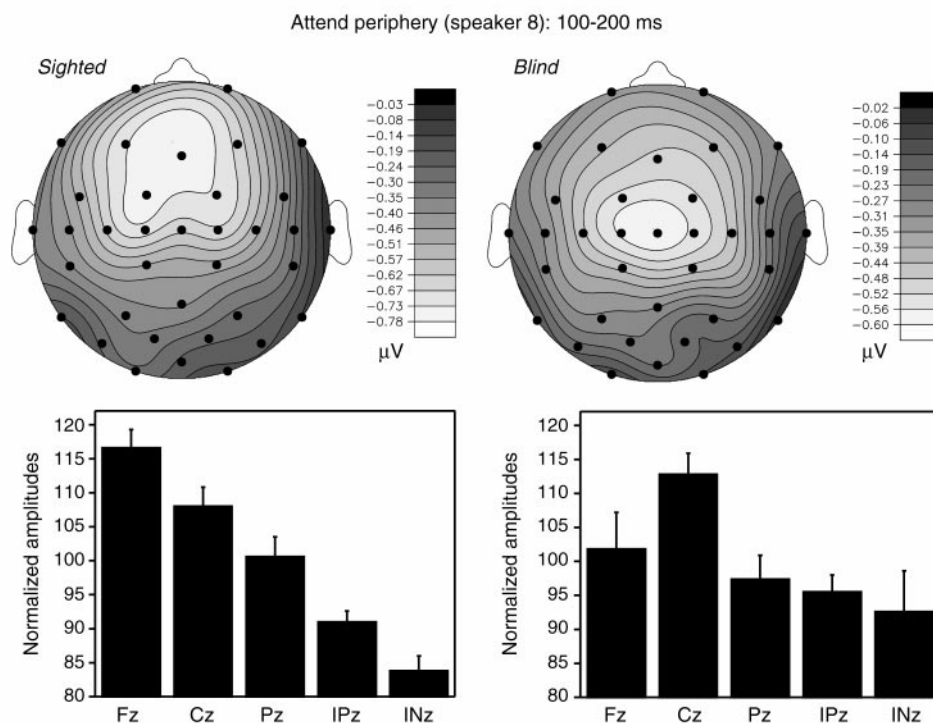


Figure 3 Topographic voltage maps of the N1 attention effect (attended minus unattended amplitudes) and the normalized anterior-posterior scalp distributions for the attended peripheral speaker. Left, sighted subjects; right, blind subjects. Lighter shading in the topographic maps indicates the greater amplitude of the N1 to attended relative to unattended standards. Bar graphs

show standardized amplitudes of the N1 attention effect (mean, 100; s.d., 15) at frontal (Fz), central (Cz), parietal (Pz), parieto-occipital (IPz) and inferior-occipital (INz) electrode sites. The anterior-posterior distribution of the N1 attention effect differed between groups with a frontal maximum in the sighted subjects and a central maximum in the blind subjects.

when participants were attending to central compared with peripheral space (see Fig. 2; speaker, $F(3, 42) = 23.19$, $P < 0.001$; condition $\times$ speaker, $F(3, 42) = 12.38$, $P < 0.001$). However, whereas the gradient of N1 amplitude did not differ significantly between the two groups when attending to the centre, during attention to the periphery the N1 gradient was significantly steeper for the blind than for the sighted individuals (condition $\times$ speaker $\times$ group, $F(3, 42) = 3.26$, $P < 0.05$; speaker $\times$ group interaction for 'attend periphery' condition, $F(3, 42) = 4.69$, $P < 0.02$; for the 'attend centre' condition, $F(3, 42) = 2.26$, $P > 0.1$).

To examine whether there is compensatory reorganization of the neural systems for early auditory selection in the blind, we compared the scalp distributions of the enhanced N1 component during auditory attention in the two groups. As shown in Fig. 3, in the sighted group the enhanced negativity was largest over the anterior scalp, whereas in the blind it was shifted posteriorly. This difference was only significant in the 'attend periphery' condition (condition $\times$ electrode (midline recordings) $\times$ group, $F(4, 56) = 3.29$, $P < 0.04$; 'attend periphery' condition, $F(4, 56) = 2.83$, $P < 0.05$; 'attend centre' condition, electrode $\times$ group, $F(4, 56) = 1.61$, $P > 0.2$).

Our behavioural results are in agreement with a growing number of studies that report that blind individuals have equal or better auditory localization than sighted individuals^{1-3,22}. Furthermore, our data show that this improved spatial acuity is not uniform but displays directional specificity: the advantage for blind subjects was found only at spatial positions where auditory localization is poorest in sighted humans, that is, at far lateral locations²¹. An improved focusing of attention in the periphery in the blind subjects was also evident in the spatial gradient of the N1 component, indicating that compensatory enhancement of early auditory spatial selection mechanisms may occur following visual deprivation from birth. Additionally, it is likely that the lack of the early neural stage of selectivity in the sighted (indexed by N1) contributed to the less finely tuned pattern of behavioural responses to peripheral sounds.

The posterior shift in the scalp topography of the enhanced N1 component in the blind subjects provides strong evidence for a reorganization of the neural substrates for early auditory selection. Previous research has reported a posterior shift of a later ERP response linked to target detection^{7,12}. Studies in visually deprived animals indicate that such alterations in topography may arise from a recruitment of posterior multimodal brain areas in which visual space is represented in sighted individuals. For example, an increase in responsiveness to auditory and/or somatosensory stimuli after visual deprivation has been reported in multimodal areas, including the superior colliculus²³, parietal cortex²⁴ and the anterior ectosylvian sulcus⁴. In addition, neurons in this last area displayed sharper auditory spatial tuning in visually deprived than in sighted cats⁴. It has also been proposed that unimodal visual areas may be recruited for non-visual processing as well when visual input is absent²⁵.

Finally, we must consider how such a refined auditory spatial representation arises in the absence of visual guidance in the congenitally blind. Developmental studies have shown that human newborns orientate to sounds even in the dark and before the emergence of visual orientating²⁶, indicating that auditory spatial representations may develop initially on the basis of sensorimotor feedback alone⁴. However, when visual input becomes available it appears to dominate and modulate the spatial representations of the remaining senses^{2,27,28}. Our findings suggest that, when visual input is congenitally absent in human development, the early auditory spatial representations continue to develop and become increasingly refined.

Our combined behavioural and ERP data show that blind individuals have an enhanced capability for sound localization in peripheral space which appears to be mediated, at least in part, by an attentional tuning mechanism that operates within the first 100 ms

after sound onset. These results are analogous to findings from studies of congenitally deaf people that report enhanced early processing of visual events in peripheral, but not central, space^{13,14}. It seems that the brain reacts to the loss of one of the two main distance senses (vision or audition) by compensatory changes in the remaining modality in a parallel and specific manner. □

Methods

Participants. Eight congenitally blind adults and eight sighted but blindfolded control subjects, matched for age, gender and handedness, participated in the study (Table 1).

Stimuli and apparatus. Acoustic stimuli (noise bursts) were delivered from an array of eight matched speakers that were mounted on a horizontally orientated metal hoop at a distance of 1.2 m from the subject's head (Fig. 1). Speakers were placed at azimuthal positions 0°, 6°, 12° and 18° (central array) and at 72°, 78°, 84° and 90° (peripheral array). Bursts of broadband pink noise (76 dB, 83 ms duration) of two different bandwidths were used as 'standards' (500–5,000 Hz, $P = 0.84$) and 'deviants' (500–15,000 Hz, $P = 0.16$), respectively. Head movements were discouraged and were monitored by means of an infrared beam reflected from a mirror attached to the electrode cap¹⁹.

Procedure. Subjects were seated in a sound-attenuated chamber throughout the experiment. Sighted participants were blindfolded. There were two separate attention conditions: subjects were instructed to attend to speaker 1 (at 0° azimuth) on half of the runs and to speaker 8 (90° azimuth) on the other half, with the task of pressing a button in response to deviants at the attended speaker (designated as 'targets') only. The responding hand was counter-balanced across subjects. Although participants were encouraged to press the button as quickly as possible, accuracy was stressed. A correct target detection was counted if the response occurred within a window of 200–800 ms after the target; a false alarm response to deviants from adjacent speakers was counted in a similar manner. Before the first experimental run, subjects were familiarized with the task and trained with one to two runs per condition. Each run (duration 2.5 min) consisted of 960 noise bursts that were presented with an average ISI of 180 ms (range 90–270 ms, rectangular distribution). A total of 10–12 runs per condition ('attend centre' or 'attend periphery') were presented.

Recording and analysis. ERPs were recorded from 41 scalp sites using tin electrodes mounted in an elastic cap (Electro-Cap International). ERPs were recorded from the following sites: frontal, Fp1, Fp2, F7, F3, Fz, F4, F8; fronto-central, FC5, FC1, FC2, FC6; temporal and central, T3, CT5, C5, C3, C1, Cz, C2, C4, C6, CT6, T4; central and parietal, P3, CP1, Pz, CP2, P4; temporo-occipital, IN5, T5, TO1, IPz, TO2, T6, IN6; parieto-occipital, PO1, PO2; and occipital, O1, O2, IN3, INz, IN4 (locations are specified in ref. 29). All recordings were referred to a right mastoid reference; an averaged left/right mastoid reference was calculated offline using an additional left mastoid recording. Vertical eye movements were monitored using an electrode below the left eye against the reference, and lateral eye movements were measured with electrodes placed at the outer canthi of the left and right eyes (bipolar recording). Electrophysiological recordings were amplified with a bandpass of 0.1–100 Hz. The electroencephalogram and electrooculogram were recorded continuously and digitized at 250 Hz. Experimental effects were analysed by analyses of variance with group (sighted versus blind) as a between-subjects factor; all other factors (condition, speaker and electrode site) were repeated-measure factors. P values were corrected according to the method of ref. 30.

Received 8 March; accepted 9 May 1999.

1. Muchnik, C., Efrati, M., Nemeth, E., Malin, M. & Hildesheimer, M. Central auditory skills in blind and sighted subjects. *Scand. Audiol.* **20**, 19–23 (1991).
2. Lessard, N., Pareé, Lepore, F. & Lassonde, M. Early-blind human subjects localize sound sources better than sighted subjects. *Nature* **395**, 278–280 (1998).
3. Rice, C. E. Early experience and perceptual enhancement. *Res. Bull. Am. Found. Blind* **22**, 1–22 (1970).
4. Rauschecker, J. P. Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* **18**, 36–43 (1995).
5. Miller, L. Diderot reconsidered: Visual impairment and auditory compensation. *J. Vis. Impair. Blind.* **86**, 206–210 (1992).
6. Röder, B., Röslér, F. & Hennighausen, E. Different cortical activation patterns in blind and sighted humans during encoding and transformation of haptic images. *Psychophysiology* **34**, 292–307 (1997).
7. Röder, B., Röslér, F., Hennighausen, E. & Näcker, F. Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cogn. Brain Res.* **4**, 77–93 (1996).
8. Röslér, F., Röder, B., Heil, M. & Hennighausen, E. Topographic differences of slow event-related brain potentials in blind and sighted adult human subjects during haptic mental rotation. *Cogn. Brain Res.* **1**, 145–159 (1993).

9. Wanet-Defalque, M. C. *et al.* High metabolic activity in the visual cortex of early blind human subjects. *Brain Res.* **446**, 369–373 (1988).
10. Sadato, N. *et al.* Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **380**, 526–528 (1996).
11. Uhl, F., Franzen, P., Podreka, I., Steiner, M. & Deecke, L. Increased regional cerebral blood flow in inferior occipital cortex and cerebellum of early blind humans. *Neurosci. Lett.* **150**, 162–164 (1993).
12. Kujala, T., Alho, K., Paavilainen, P., Summala, H. & Näätänen, R. Neuronal plasticity in processing of sound location by the early blind: an event-related potential study. *Electroencephalogr. Clin. Neurophysiol.* **84**, 469–472 (1992).
13. Neville, H. J. & Lawson, D. Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Res.* **405**, 268–283 (1987).
14. Neville, H. J., Schmidt, A. & Kutas, M. Altered visual-evoked potentials in congenitally deaf adults. *Brain Res.* **266**, 127–132 (1983).
15. Downing, C. J. & Pinker, S. in *Attention and Performance* Vol. XI (eds Posner, M. & Martin, O. S. M.) 171–188 (Erlbaum, Hillsdale, New Jersey, 1985).
16. Mondor, T. A. & Zatorre, R. J. Shifting and focusing auditory spatial attention. *J. Exp. Psychol. Hum. Percept.* **21**, 387–409 (1995).
17. Hillyard, S. A., Mangun, G. R., Woldorff, M. C. & Luck, S. J. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 665–681 (MIT Press, Cambridge, Massachusetts, 1995).
18. Näätänen, R. *Attention and Brain Function* (Erlbaum, Hillsdale, New Jersey, 1992).
19. Teder-Sälejärvi, W. & Hillyard, S. A. The gradient of spatial auditory attention in free field: An event-related potential study. *Percept. Psychophys.* **60**, 1228–1242 (1998).
20. Middlebrooks, J. C. & Green, D. M. Sound localization by human listeners. *Annu. Rev. Psychol.* **42**, 135–159 (1991).
21. Mills, A. in *Foundations of Modern Auditory Theory* (ed. Fobias, J. V.) 303–348 (Academic, New York, 1972).
22. Ashmead, D. H. *et al.* Spatial hearing in blind children with visual disabilities. *Perception* **27**, 105–122 (1998).
23. Rauschecker, J. P. & Harris, L. R. Auditory compensation of the effect of visual deprivation in cats' superior colliculus. *Exp. Brain Res.* **50**, 63–83 (1983).
24. Hyvärinen, J., Hyvärinen, L. & Linnankoski, I. Modification of parietal association cortex and functional blindness after binocular deprivation in young monkeys. *Exp. Brain Res.* **42**, 1–8 (1981).
25. Cohen, L. G. *et al.* Functional relevance of cross-modal plasticity in blind humans. *Nature* **389**, 180–183 (1997).
26. Morrongiello, B. A. in *The Development of Intersensory Perception* (eds Lewkowicz, D. J. & Lickliter, R.) 235–263 (Erlbaum, Hillsdale, New Jersey, 1994).
27. Brainard, M. S. & Knudsen, E. I. Sensitive periods for visual calibration of auditory space map in the barn owl optic tectum. *J. Neurosci.* **10**, 3929–3942 (1998).
28. Knudsen, E. I. Capacity for plasticity in the adult owl auditory system expanded by juvenile experience. *Science* **279**, 1531–1533 (1998).
29. Clark, V. P., Fan, S. & Hillyard, S. A. Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Hum. Brain. Map.* **2**, 170–187 (1995).
30. Huynh, H. & Feldt, L. S. Estimation of the box correction for degrees of freedom from sample data in randomized block and splitplot designs. *J. Edu. Stat.* **1**, 69–82 (1976).

Acknowledgements. We thank the San Diego Center for the Blind and Visually Impaired, the Braille Institute of America (San Diego), and the Blind Recreation Center in San Diego for help in recruiting blind participants; L. Anillo-Vento for helping to recruit matched sighted controls; C. Nava for help during data acquisition; and M. Marlow for computer assistance. This work was supported by the German Research Foundation (DFG) and grants from the National Institutes of Mental Health and National Institutes of Health.

Correspondence and requests for material should be addressed to B.R. (e-mail: roeder@mail.uni-marburg.de).

Distortion of proximodistal information causes JNK-dependent apoptosis in *Drosophila* wing

Takashi Adachi-Yamada, Konomi Fujimura-Kamada, Yasuyoshi Nishida & Kunihiro Matsumoto

Division of Biological Science, Graduate School of Science, Nagoya University, and CREST, Japan Science and Technology Corporation, Chikusa-ku, Nagoya 464-8602, Japan

Distinct and evolutionarily conserved signal-transduction cascades mediate the survival or death of cells during development. The c-Jun amino-terminal kinases (JNKs) of the mitogen-activated protein kinase superfamily are involved in apoptotic signalling in various cultured cells¹. However, the role of the JNK pathway in development is less well understood. In *Drosophila*, Decapentaplegic (Dpp; a homologue of transforming growth factor-β) and Wingless (Wg; a Wnt homologue) proteins are secretory morphogens that act cooperatively to induce formation of the proximodistal axis of appendages^{2–7}. Here we show that either decreased Dpp signalling in the distal wing cells or increased Dpp signalling in

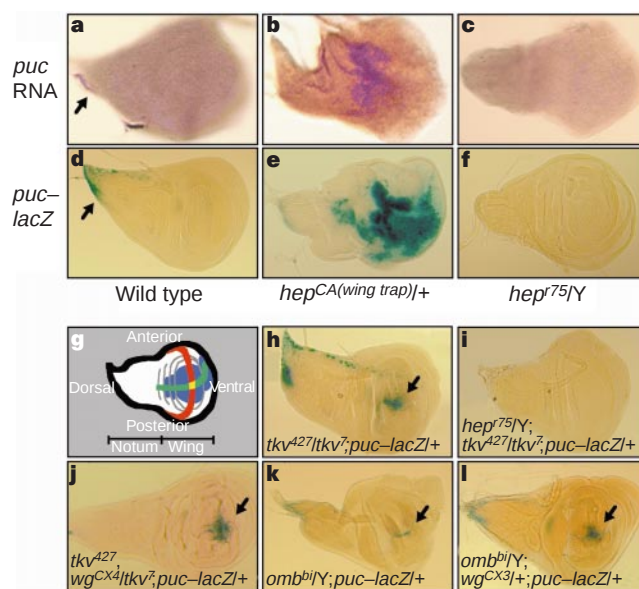


Figure 1 Interaction of the DJNK and Dpp/Wg-omb pathways in the wing disc. *puc* RNA expression is revealed by *in situ* hybridization (a–c) and enhancer trap *lacZ* reporter *E69* (ref. 15) (d–f, h–l). a, d, Wild type (a, Canton-S; d, *puc*^{E69}/+). Arrows indicate the scutellum anlage. b, e, *Hep*^{CA} producer (b, *hep*^{CA(wing trap)}/+; e, *hep*^{CA(wing trap)}/*puc*^{E69}/+). c, f, *hep* null mutant (c, *hep*⁷⁵/Y; f, *hep*⁷⁵/Y;*puc*^{E69}/+). g, Schematic representation of the wing-disc domains expressing *dpp* (green), *wg* (red) and *omb* (blue). Only the expressing domain in the wing primordia is shown for each. The area where the Dpp- and Wg-expressing domains intersect (yellow) corresponds to the distal-tip primordium. h, *tkv* hypomorph (*tkv*⁷⁵/*tkv*⁴²⁷; *puc*^{E69}/+). i, *tkv* hypomorph in a *hep* null background (*hep*⁷⁵/Y; *tkv*⁷⁵/*tkv*⁴²⁷; *puc*^{E69}/+). j, *tkv* hypomorph heterozygous for *wg* (*tkv*⁷⁵/*tkv*⁴²⁷ *wg*^{CX4}; *puc*^{E69}/+). k, *omb* hypomorph (*omb*^{bi}/Y; *puc*^{E69}/+). l, *omb* hypomorph heterozygous for *wg* (*omb*^{bi}/Y; *wg*^{CX3}/+; *puc*^{E69}/+). Arrows indicate ectopic *puc* expression in the wing-tip primordium.

the proximal wing cells causes apoptosis. Inappropriate levels of Dpp signalling lead to aberrant morphogenesis in the respective wing zones, and these apoptotic zones are also determined by the strength of the Wg signal. Our results indicate that distortion of the positional information determined by Dpp and Wg signalling gradients leads to activation of the JNK apoptotic pathway, and the consequent induction of cell death thereby maintains normal morphogenesis.

Drosophila has a JNK signalling cascade consisting of *Drosophila* JNK (DJNK) and the DJNK kinase Hep. Genetic studies have indicated that the Hep–DJNK pathway functions to promote embryonic dorsal closure by maintaining production of Dpp^{1,8,9}. Disruption of the Dpp and DJNK pathways have similar phenotypic effects in both adult development and the embryonic dorsal closure. For example, viable loss-of-function mutant alleles of *hep*, *dpp* or *thick veins* (*tkv* encoding a type-I Dpp receptor), show a similar notum cleft phenotype^{10–12}. Double mutants of the *hep* and *dpp* or *tkv* alleles resulted in a fully penetrant lethal phenotype (data not shown), indicating that they have a functional relationship extending beyond mere phenotypic similarity. To examine this relationship further, we first investigated *dpp* expression in *hep* mutants. *dpp* expression during embryonic dorsal closure is lower in *hep* mutants than in wild-type embryos^{1,8,9}. In contrast, *dpp* expression in the wing imaginal disc of the *hep* null mutant was similar to that of the wild type (data not shown), indicating that the Hep–DJNK pathway does not regulate *dpp* expression in the wing disc. Similarly, loss of Hep did not affect the expression of *wg* (data not shown), which is expressed along the dorsoventral boundary.

Puckered (Puc) is a dual-specificity phosphatase, the expression of which is induced by the DJNK pathway and which inactivates

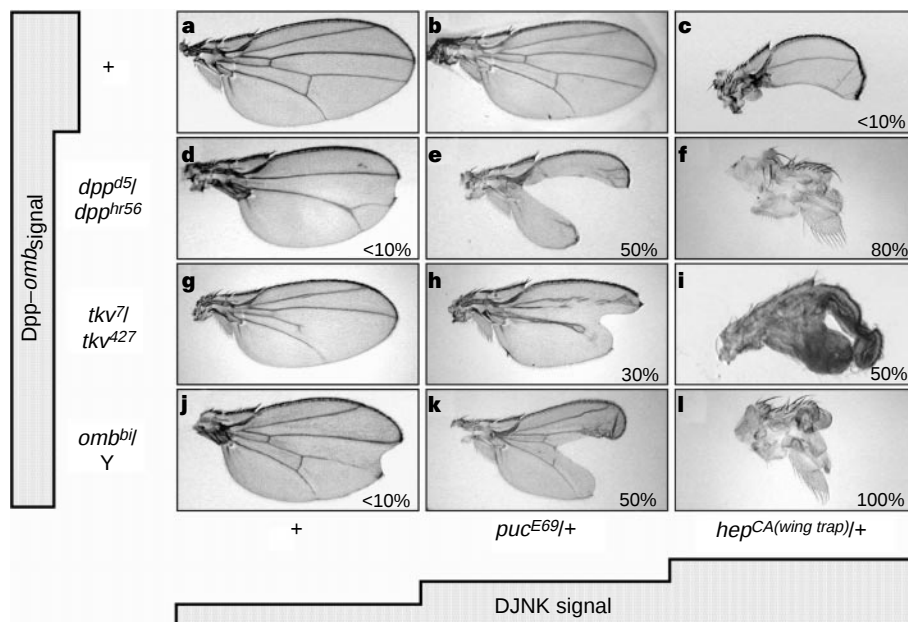


Figure 2 Interaction of the DJNK and Dpp pathways in wing formation. Percentages shown in each photograph refer to the percentage penetrance of the observed notching. **a**, Wild type (*Canton-S*). **b**, Heterozygote for *puc*. **c**, *Hep^{CA}* producer. **d**, *dpp* hypomorph. **e**, *dpp* hypomorph that is heterozygous for *puc*. **f**, *dpp* hypomorph producing *Hep^{CA}*. **g**, *tkv* hypomorph. **h**, *tkv* hypomorph that is

heterozygous for *puc*. **i**, *tkv* hypomorph producing *Hep^{CA}*. Because these flies died at the pharate adult stage, the wings were dissected out. **j**, *omb* hypomorph. **k**, *omb* hypomorph that is heterozygous for *puc*. **l**, *omb* hypomorph producing *Hep^{CA}*. Magnification in **f** and **i** is two times greater than that in other photographs.

DJNK during embryonic dorsal closure¹³. At the late-third-instar larval stage, *puc-lacZ* expression was *Hep*-dependent in most tissues except for the central nervous system (data not shown), which suggests that *puc-lacZ* expression can be used to monitor the extent of activation of the *Hep*-DJNK pathway in the imaginal discs. In the wild-type wing disc, *puc* expression was not observed in the wing blade primordium (Fig. 1a, d). When a constitutively active *hep* mutant (*hep^{CA}*) was expressed in the wing disc, *puc* was expressed in a wide area of the wing primordium (Fig. 1b, e). These results indicate that the *Hep*-DJNK pathway is inactive in the wing blade primordium of wild-type flies. Furthermore, *puc* expression detected by *in situ* hybridization was correlated with expression of *puc-lacZ* (Fig. 1a-f). Therefore, in the following studies, *puc-lacZ* lines were used to analyse *puc* expression.

To examine the relationship between the *Hep*-DJNK and Dpp-Tkv pathways, we tested the effect of mutants in the Dpp-Tkv pathway on the expression of *puc-lacZ*. In a *tkv*-mutant fly, *puc-lacZ* was expressed at the primordial wing tip in a *Hep*-dependent manner (Fig. 1h, i). Similar induction of *puc-lacZ* was also observed in a *dpp* mutant, *dpp^{d8}/dpp^{hr56}* (data not shown). These results indicate that the Tkv pathway negatively regulates the DJNK pathway. The domain expressing *puc-lacZ* forms an island on the primordial distal tip where the anteroposterior and dorsoventral axes intersect (Fig. 1g). This suggests that some additional signalling, implicated in the formation of dorsoventral polarity, is involved in the regulation of *puc-lacZ* expression. One such candidate is signalling from Wg. Consistent with this possibility, the expression of *puc-lacZ* in the primordial distal tip was enhanced when the fly was made heterozygous for *wg* (Fig. 1j). Similarly, elevated expression of *puc-lacZ* was observed in a fly containing a mutant allele of *optomotor-blind* (*omb*) (Fig. 1k), a gene encoding a T-box transcription factor which is positively regulated by both Dpp and Wg in wing discs¹⁴ (Fig. 1g). Furthermore, this *puc* induction by the *omb* mutation was further enhanced by reducing the gene dosage of *wg* (Fig. 1l). Taken together, these results indicate that signalling from Dpp and Wg to the *omb* gene downregulates the DJNK pathway.

These results led us to expect that further activation of the *Hep*-DJNK pathway in mutants of the Dpp-*omb* pathway would result in

the enhancement of the phenotype with regard to wing formation. Activation of the DJNK pathway was increased when the gene dosage of *puc* was reduced in heterozygote flies or by introducing *hep^{CA}* (Fig. 2). In mutants of *dpp*, *tkv* or *omb*, a fraction of adults showed characteristic small notching in the wing tip (Fig. 2d, g, j). When the dosage of *puc* was reduced in the mutant background, this notching phenotype was significantly enhanced (Fig. 2e, h, k), whereas the same heterozygote gene dose of *puc* in a wild-type background had no phenotype (Fig. 2b). Furthermore, introduction of *hep^{CA}* into these mutants resulted in the loss of a wide region of the wing blade, leaving only the base of wing (Fig. 2f, i, l). These results indicate that decreased Dpp-*omb* signalling activates the *Hep*-DJNK pathway, which leads to loss of wing formation.

We next investigated cell death in wing discs by staining with acridine orange. In the late-third-instar larva, few dying cells were observed in the wild-type wing blade primordium (Fig. 3a), as reported previously¹⁵. In a *puc* heterozygote, the number of apoptotic cells did not increase (Fig. 3b). This lack of effect is consistent with the observation that *puc* is not expressed in the wild-type wing primordium (Fig. 1a, d). In a *tkv* mutant, a cluster of dying cells appeared in the primordial wing tip (Fig. 3c), where the DJNK pathway is activated (Fig. 1h). The appearance of this apoptotic cluster was more evident in a heterozygous *puc* background, where *puc* gene dosage is reduced (Fig. 3d). A *hep* null allele suppressed the appearance of apoptotic cell clusters in the primordial wing tip of the *tkv* mutant (Fig. 3e), indicating that this cell death is dependent on the *Hep*-DJNK pathway. Furthermore, massive cell death was observed in both the proximal and distal wing primordia in the *hep^{CA}* producer (Fig. 3f). Thus, activation of the *Hep*-DJNK pathway is sufficient for cell death within a wide area of the wing primordium.

An apoptotic cluster was also observed in the wing-tip primordium of the *omb* mutant (Fig. 3g) and this apoptotic cluster appeared more conspicuously in an *omb* mutant that was heterozygous for *wg* (Fig. 3h). These results are consistent with the observation that the DJNK pathway is activated in these mutants (Fig. 1k, l). Therefore, it is likely that the Wg signal functions as a survival signal by inducing *omb* expression in the distal wing. The *Drosophila* Ras-Raf-ERK signal has been shown to promote cell survival by downregulating

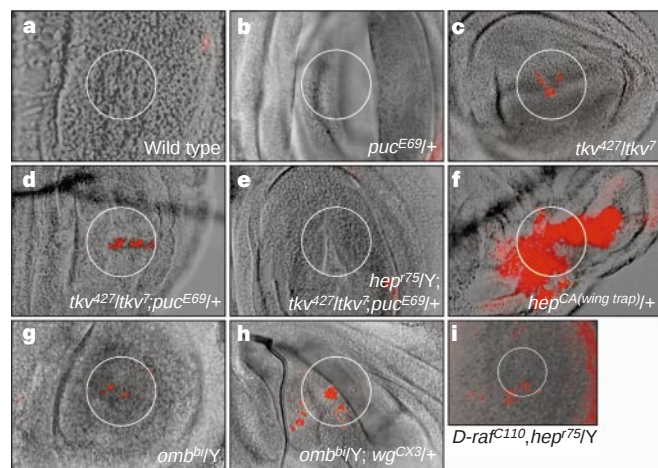


Figure 3 Cell death in the wing disc. All photographs were made by superimposing the fluorescent image generated by acridine-orange staining (red) with the visual light image. Circles indicate the location of the wing-tip primordium. **a**, Wild type (*Canton-S*). (Number of dying cells at the distal-tip primordium ($\pm$ standard deviation), 0.2 ± 0.6). **b**, Heterozygote for *puc* (0.1 ± 0.3). **c**, *tkv* hypomorph (5.1 ± 3.5). **d**, *tkv* hypomorph that is heterozygous for *puc* (13.2 ± 8.2). **e**, *tkv* hypomorph that is heterozygous for *puc* in a *hep* null background (0.0 ± 0.0). **f**, *Hep^{CA}* producer (always more than 100, not countable). **g**, *omb* hypomorph (3.8 ± 4.1). **h**, *omb* hypomorph that is heterozygous for *wg* (9.3 ± 3.2). **i**, *D-raf* hypomorph in a *hep* null background. The cell number and the size of the wing disc of this genotype are smaller than those of wild type. Thus the same magnification of the corresponding wing regions results in the different size of photos. At least 40 wing discs were examined in each experiment except for the experiment shown in **e**. In **e**, four discs were examined. More than 80% of samples showed the observed phenotypes.

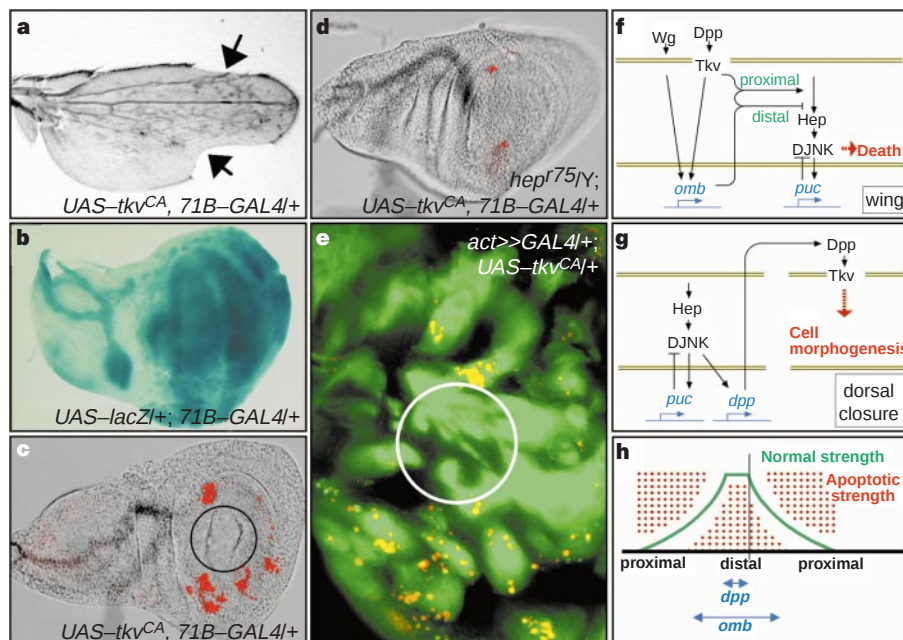


Figure 4 Increased Dpp signalling induces apoptosis in the proximal wing. **a**, A wing expressing *Tkv^{CA}*. Arrows indicate notches located away from the distal tip. **b**, Expression pattern of *71B-GAL4*. **c**, **d**, Acridine-orange staining of the wing discs. *Tkv^{CA}* is induced by *71B-GAL4* in a *hep⁺* (**c**) and a *hep* null (**d**) background. Circle indicates the location of the distal-wing primordium. **e**, Acridine orange staining of the wing disc with *Tkv^{CA}*-expressing clones (*hsFlp/Y; AyGAL4, UAS-GFP/+; UAS-tkv^{CA}/+*) (green). Circle indicates the location of the primordial distal tip. Red and yellow indicate dying cells. **f**, A proposed pathway in which Dpp and Wg signals inhibit DJNK-dependent apoptosis in the distal wing cells, but stimulate it in the proximal cells. **g**, During embryonic dorsal closure, DJNK

signalling regulates Dpp production to elicit elongation of adjacent cells^{18,9}. **h**, A model for apoptosis induction. Apoptosis is determined by the proximodistal positional values along a cross section of the primordial wing tip. Green line, the strength of Dpp signal along the proximodistal axis of normal flies^{4,5,7}; vertical line, the boundary of the anteroposterior compartment; blue arrow, the approximate positions of the Dpp source and the *omb*-expressing domain; red stippled area, the strength and position along the proximodistal axis of the signal predicted to induce apoptosis. When the green line enters in the red stippled area, the gradient of signalling strength is distorted, resulting in apoptosis.

Hid function and expression^{16,17}. A loss-of-function mutation in *Drosophila raf* (*D-raf*) also induced massive apoptosis throughout the primordial wing blade (data not shown). However, this apoptosis occurred in a *Hep*-independent manner (Fig. 3i). Thus, the *Hep*-DJNK pathway is specifically involved in apoptosis mediated by the Dpp/Wg-*omb* pathway. These observations are consistent with the following model of proximodistal axis formation^{2,7}: Dpp and Wg are produced along the anteroposterior and dorsoventral boundaries, respectively (Fig. 1g), and their intersecting gradients provide the positional information that determines the location of various wing subdomains.

Bone morphogenetic protein (BMP)-2 and -4, which are vertebrate homologues of Dpp, have been shown to induce apoptosis¹⁸. We therefore examined whether upregulation of the Dpp pathway in the wing disc also causes apoptosis. When a constitutively active *Tkv* (*Tkv^{CA}*)¹⁹ was induced in the wing disc by the *GAL4/UAS* system using the *GAL4* enhancer trap line *71B* (ref. 20), the wing margin was frequently notched (Fig. 4a). Although expression of *GAL4* was higher around the wing tip than in the proximal region (Fig. 4b), the notches were always positioned away from the distal tip and were limited to the more proximal region. Staining with acridine orange also revealed that dying cells were distributed through the primordium of the proximal wing but were absent in the distal wing primordium (Fig. 4d), indicating that apoptosis in the proximal wing region is dependent on the DJNK pathway. When *Tkv^{CA}*-expressing clones were generated mosaically in the wing primordia, dying cells were detected within the *Tkv^{CA}* clones in the proximal wing but not in the distal wing (Fig. 4e). Therefore, upregulation of Dpp signalling can induce apoptosis in the proximal cells, but not in the distal cells (Fig. 4f, h).

Our results indicate that upregulation of the Dpp signal in the proximal wing or downregulation of the Dpp signal in the distal

wing causes DJNK-mediated apoptosis (Fig. 4f, h). Because aberrant Dpp signalling in either case is deleterious for normal wing morphogenesis, ablation of cells by apoptosis may be an effective mechanism to restore and maintain the normal process of wing development. Downregulation of the DJNK pathway by Dpp signalling is also essential for distal-tip cell survival in several other appendages in addition to the wing. Loss of distal regions owing to cell death in appendages other than the wing have also been observed in *Drosophila* carrying some alleles of *dpp* or *wg*^{21,22}. These phenotypes were enhanced in flies with a reduced dosage of the *puc* gene (data not shown), similar to the phenotype we observed in the wing. On the basis of these results, we propose that the regulation of DJNK-dependent apoptosis by a signal gradient providing proximodistal positional value is a general mechanism in appendage development, and is not restricted to the wing. We believe that the DJNK pathway is required for restoration of proximodistal positional information only when signalling levels are abnormal and that normal wing development may not require DJNK signalling. Thus, the DJNK cascade is latent in normal development, but is activated to provide a system to maintain proper development when normal signalling is distorted. □

Methods

A mutant strain *tkv*⁴²⁷. *tkv*⁴²⁷ was once reported to be a hyperactive allele based on the result that its wing-vein phenotype was enhanced by an additional *dpp*⁺ transgene²³. However, similar phenomena can be observed not only in *tkv*⁴²⁷ but also in a wide variety of loss-of-function *tkv* alleles (data not shown). Recently, it was also reported that *Tkv* represses the expression of Dpp required for vein formation during the pupal stage²⁴. Thus, hypomorphs of *tkv* produce abundant amounts of Dpp and show the 'thick veins' phenotype enhanced by *dpp*⁺ transgene.

Construction of a transgenic fly *hep*^{CA(wing trap)}. MKK7, a mammalian homologue of Hep, is activated by phosphorylation at Ser271, Thr275 and Ser277 (ref. 25). Hence, the corresponding amino-acid residues, Ser346, Thr350 and Ser352 in Hep were all replaced with Asp to generate Hep^{CA}. Among 20 strains of UAS-*hep*^{CA} transgenic flies, one insertion in the third chromosome manifested the wing-notch phenotype without GAL4 drivers with a penetrance of less than 10% (Fig. 2c). Furthermore, because ectopic *puc-lacZ* was expressed in the wing disc of this strain (Fig. 1b), the wing-notch phenotype was presumed to be caused by the enhancer trap of the UAS-*hep*^{CA} transgene. This strain was named *hep*^{CA(wing trap)} and used throughout this work.

Acridine-orange staining of the wing disc. Staining was carried out as described²⁶. Although apoptotic cell clusters were observed in various genetic backgrounds, these were usually accompanied by somewhat elevated levels of sporadic dying cells throughout the wing disc. It has been reported previously that apoptosis induced in a restricted wing area is typically accompanied by nonautonomous death in adjacent territories¹⁵.

Clonal analysis of *Tkv*^{CA}-induced cell death in the *AyGAL4* system. The *AyGAL4* system is a transgene fusing the following DNA segments: *Actin5C* promoter-FRT (yeast FLP recombinase target)-transcriptional termination signal-yellow⁺-FRT-GAL4 (ref. 27). Discs were prepared from larvae carrying both the *AyGAL4* and *hsFlp* (yeast FLP recombinase gene driven by a heat shock promoter) transgenes. GAL4-expressing clones were induced by heat treatment at 37 °C for 30 min at 48–72 h after egg laying. Staining was carried out at 48 h after heat treatment.

Received 5 January; accepted 20 May 1999.

- Ip, Y. T. & Davis, R. J. Signal transduction by the c-Jun N-terminal kinase (JNK)—from inflammation to development. *Curr. Opin. Cell Biol.* **10**, 205–219 (1998).
- Campbell, G., Weaver, T. & Tomlinson, A. Axis specification in the developing *Drosophila* appendage: the role of *wingless*, *decapentaplegic*, and the homeobox gene *aristless*. *Cell* **74**, 1113–1123 (1993).
- Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of *Wingless* morphogen gradient. *Cell* **87**, 833–844 (1996).
- Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of DPP morphogen gradient. *Cell* **85**, 357–368 (1996).
- Lecuit, T. et al. Two distinct mechanisms for long-range patterning by *Decapentaplegic* in the *Drosophila* wing. *Nature* **381**, 387–393 (1996).
- Neumann, C. J. & Cohen, S. Long-range action of *Wingless* organizes the dorsal-ventral axis of the *Drosophila* wing. *Development*, **124**, 871–880 (1997).
- Neumann, C. J. & Cohen, S. Morphogens and pattern formation. *BioEssays* **19**, 721–729 (1997).
- Noselli, J. JNK signalling and morphogenesis in *Drosophila*. *Trends Genet.* **14**, 33–38 (1998).

- Goberdohan, D. C. I. & Wilson, C. JNK, cytoskeletal regulator and stress response kinase? A *Drosophila* perspective. *BioEssays* **20**, 1009–1019 (1998).
- Segal, D. & Gelbart, W. M. Shortvein, a new component of the decapentaplegic gene complex in *Drosophila melanogaster*. *Genetics* **109**, 119–143 (1985).
- Terracol, R. & Lengyel, J. A. The thick veins gene of *Drosophila* is required for dorsoventral polarity of the embryo. *Genetics* **138**, 165–178 (1994).
- Riesgo-Escovar, J. R. & Hafen, E. Common and distinct roles of Dfos and Djun during *Drosophila* development. *Science* **278**, 669–672 (1997).
- Martin-Blanco, E. et al. *puckered* encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in *Drosophila*. *Genes Dev.* **12**, 557–570 (1998).
- Grimm, S. & Pflugfelder, G. O. Control of the gene *optomotor-blind* in *Drosophila* wing development by *decapentaplegic* and *wingless*. *Science* **271**, 1601–1604 (1996).
- Milán, M., Campuzano, S. & García-Bellido, A. Developmental parameters of cell death in the wing disc of *Drosophila*. *Proc. Natl Acad. Sci. USA* **94**, 5691–5696 (1997).
- Kurada, P. & White, K. Ras promotes cell survival in *Drosophila* by downregulating *hid* expression. *Cell* **95**, 319–329 (1998).
- Bergmann, A., Agapite, J., McCall, K. & Steller, H. The *Drosophila* gene *hid* is a direct molecular target of Ras-dependent survival signaling. *Cell* **95**, 331–341 (1998).
- Hogan, B. L. M. Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev.* **10**, 1580–1594 (1996).
- Adachi-Yamada, T. et al. p38 mitogen-activated protein kinase can be involved in transforming growth factor-β superfamily signal transduction in *Drosophila* wing morphogenesis. *Mol. Cell. Biol.* **19**, 2322–2329 (1999).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotype. *Development* **118**, 401–415 (1993).
- Spencer, F. A., Hoffman, F. M. & Gelbart, W. M. Decapentaplegic: a gene complex affecting morphogenesis in *Drosophila melanogaster*. *Cell* **28**, 451–461 (1982).
- Couso, J. P., Bate, M. & Martínez-Arias, A. A *wingless*-dependent polar coordinate system in *Drosophila* imaginal discs. *Science* **259**, 484–489 (1993).
- Penton, A. et al. Identification of two bone morphogenetic protein type I receptors in *Drosophila* and evidence that Brk25D is a *decapentaplegic* receptor. *Cell* **78**, 239–250 (1994).
- de Celis, J. F. Expression and function of *decapentaplegic* and *thick veins* during the differentiation of the veins in the *Drosophila* wing. *Development* **124**, 1007–1018 (1997).
- Holland, P. M., Suzanne, M., Campbell, J. S., Noselli, S. & Cooper, J. A. MKK7 is a stress-activated mitogen-activated protein kinase functionally related to *hemipterous*. *J. Biol. Chem.* **272**, 24994–24998 (1997).
- Masucci, J. D., Miltenberger, R. J. & Hoffmann, F. M. Pattern-specific expression of the *Drosophila* *decapentaplegic* gene in imaginal discs is regulated by 3'-cis regulatory elements. *Genes Dev.* **4**, 2011–2023 (1990).
- Ito, K., Awano, W., Suzuki, K., Hiromi, Y. & Yamamoto, D. The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurons and glial cells. *Development* **124**, 761–771 (1997).

Acknowledgements. We thank K. Basler, K. Ekstrom, W. M. Gelbart, Y. Hiromi, Y. H. Inoue, T. Kadowaki, K. Matthews, M. Nakamura, S. Noselli, H. Okano, N. Perrimon, G. O. Pflugfelder, G. Struhl, K. Takahashi, D. Yamamoto and S. Yoshikawa for materials; and K. W. Cho, S. Goto, A. Kuroiwa, M. Lamphier, E. Nishida, M. B. O'Connor, T. Tabata and Y. Tomoyasu for advice on the manuscript. Supported by grants from The Award of Tokai Scholarship Foundation, The Kurata Foundation, The Ministry of Education, Science, Sports and Culture of Japan, and Japan Science and Technology Corporation.

Correspondence and requests for materials should be addressed to T.A.-Y. (e-mail: adachi@bio.nagoya-u.ac.jp).

The *mPer2* gene encodes a functional component of the mammalian circadian clock

Binhai Zheng*, David W. Larkin†, Urs Albrecht‡, Zhong Sheng Sun*, Marijke Sage*, Gregor Eichele‡, Cheng Chi Lee* & Allan Bradley*§

* Department of Molecular and Human Genetics; † Division of Neuroscience; ‡ Verna and Marrs McLean Department of Biochemistry; and § Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

Circadian rhythms are driven by endogenous biological clocks that regulate many biochemical, physiological and behavioural processes in a wide range of life forms¹. In mammals, there is a master circadian clock in the suprachiasmatic nucleus of the anterior hypothalamus. Three putative mammalian homologues (*mPer1*, *mPer2* and *mPer3*) of the *Drosophila* circadian clock gene *period* (*per*) have been identified^{2–4}. The *mPer* genes share a conserved PAS domain (a dimerization domain found in *Per*, *Arnt* and *Sim*) and show a circadian expression pattern in the suprachiasmatic nucleus. To assess the *in vivo* function of *mPer2*, we generated and characterized a deletion mutation in the PAS domain of the mouse *mPer2* gene. Here we show that mice

|| Present address: Max Planck Institute for Experimental Endocrinology, Feodor-Lynen-Strasse 7, 30625 Hanover, Germany.

homozygous for this mutation display a shorter circadian period followed by a loss of circadian rhythmicity in constant darkness. The mutation also diminishes the oscillating expression of both *mPer1* and *mPer2* in the suprachiasmatic nucleus, indicating that *mPer2* may regulate *mPer1* *in vivo*. These data provide evidence that an *mPer* gene functions in the circadian clock, and define *mPer2* as a component of the mammalian circadian oscillator.

The PAS domain is highly conserved in circadian clock genes from a variety of species^{9–13}. We designed a replacement vector to delete two exons encoding the most conserved region of *mPer2* in comparison with *Drosophila* *Per* (Fig. 1a). This deletion removes half of PAS B and the entire PAC subdomain¹⁴ and is expected to disrupt the PAS domain. The targeted allele (*mPer2*^{Brdm1}, abbreviated as m) was obtained in AB2.2 embryonic stem cells (derived from an XY 129S7/SvEvBrd-Hprt^{b-m2} embryo, abbreviated as 129 below) (Fig. 1b) and used to generate germline chimaeric mice. Intercrosses between heterozygous (C57BL/6 × 129) F₁ offspring produced homozygous mutants (Fig. 1c) at the expected Mendelian ratio. Homozygous mutants (m/m) were morphologically indistinguishable from their wild-type littermates and both males and females were fertile. Reverse transcription with polymerase chain reaction (RT-PCR) and sequence analysis indicated the presence of a mutant transcript that, if translated, would generate a protein with a deletion of 87 amino acids (Fig. 1d).

To determine whether the mutation altered circadian behaviour,

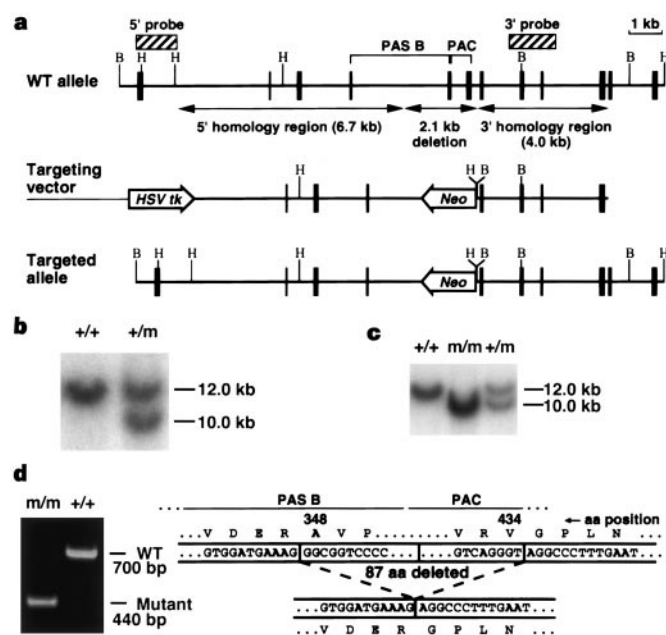


Figure 1 Generation of *mPer2*^{Brdm1} mutant mice. **a**, Genomic structure of a portion of the mouse *mPer2* gene, the targeting vector and the predicted structure of the targeted allele. Exons are indicated by vertical black bars. WT, wild type; H, *HindIII*; B, *BamHI*; *Neo*, neomycin resistance gene; *HSV tk*, Herpes simplex virus thymidine kinase gene. **b**, Southern analysis to identify targeted embryonic stem cell clones by digesting genomic DNA with *BamHI*. A 1.2-kb 5' external probe detects a 12.0-kb wild-type (+) fragment and a 10.0-kb mutant (m) fragment. A 1.4-kb 3' internal probe that detects an 11.2-kb *HindIII* wild-type fragment and a 5.7-kb mutant fragment was used to confirm correct targeting at the 3' homology region (data not shown). **c**, Southern analysis of F₂ littermates obtained from intercrosses between (C57BL/6 × 129) F₁ heterozygous mice using the same restriction enzyme and probe as in **b**. A probe corresponding to the deleted region was used in Southern analysis to confirm the deletion in homozygous mutants (data not shown). **d**, RT-PCR and sequence of the mutant transcripts. Left: RT-PCR products separated on an agarose gel which were amplified with cDNA primers flanking the expected deletion. bp, base pair. Right: comparison of the wild-type and mutant cDNA sequences with conceptual translation shown. Amino-acid position is from GenBank accession number AF036893.

F₂ homozygous mutants and their wild-type and heterozygous littermates were individually housed in circadian activity-monitoring chambers. Wheel-running activity, which is an accurate measure of circadian activity in rodents, was monitored for each animal. Mice were initially maintained in a 12 h light/12 h dark (LD 12:12, or LD) cycle for 10 days to establish entrainment, and were subsequently maintained in constant darkness (DD) (Fig. 2). After 22 days in darkness, the animals were exposed to a 6-h light pulse (arrows in Fig. 2) and were subsequently maintained in darkness for an additional 11 days. F₂ wild-type mice showed a precise entrainment to the LD cycle and in constant darkness they established a circadian rhythm with a period slightly less than 24 h (Fig. 2a). The 6-h light pulse reset the clock of these animals when applied in the subjective night.

Homozygous mutants were also entrained in LD (see below). However, they displayed two overt phenotypes in constant darkness. The first phenotype of homozygous mutants is a short circadian period. At the beginning of DD (Fig. 2b) or after the light pulse (Fig. 2c), homozygous mice typically displayed a short period for several days. The period length was calculated by χ^2 periodogram analyses on intervals when the circadian period appeared stable on the activity record. Wild-type controls had an average period of 23.7 ± 0.4 h (mean $\pm$ s.d., $n = 13$). In contrast, m/m mice had an average period of 22.1 ± 0.4 h ($n = 17$), which is 4 s.d. from the mean of the wild-type controls. In some cases the precision of this rhythm in m/m mice could be visualized by replotting the activity record on a scale corresponding to the period (Fig. 2d). The stable period for heterozygous mutants averaged 23.6 ± 0.3 h ($n = 13$), which is not significantly different from that of their wild-type littermates. Thus the homozygous mutants displayed a significantly shorter period than that of their wild-type siblings ($P < 0.001$, Student's *t* test).

The second phenotype of homozygous mutants is a loss of persistent circadian rhythmicity in constant darkness. The short period expressed by homozygous mutants was followed by a loss of the circadian rhythm (Fig. 2b, c). The time that it took an individual homozygous mutant to lose its circadian rhythm varied, mostly ranging between 2 and 18 days as assessed by activity plots ($n = 17$). None of the wild-type littermates ($n = 13$) became arrhythmic during the corresponding period. Of the heterozygous animals tested ($n = 13$), two displayed multiple rhythms or transient loss of circadian rhythmicity, whereas the other heterozygous mutants were not detectably different from the wild-type controls (data not shown). To quantify circadian rhythmicity, we carried out Fourier periodogram analysis¹⁵ on days 1–10 (the first interval) and days 6–15 (the second interval) in constant darkness for the three genotypes. The logarithm of the amplitude of the circadian peak for the first and second interval averages 2.72 ± 0.14 (mean $\pm$ s.e.m.) and 2.80 ± 0.17 , respectively, for +/+ mice ($n = 13$), 2.85 ± 0.08 and 2.88 ± 0.07 , respectively, for +/m mice ($n = 13$), and 2.19 ± 0.30 and 1.80 ± 0.35 , respectively, for m/m mice ($n = 17$) ($*P < 0.05$, Student's *t* test). Fourier analysis on the activity data from homozygous mutants that have no apparent circadian rhythms confirmed the loss of the circadian peak but detected a periodicity of 4–8 h (Fig. 3). Such activity rhythms with a period less than 24 h, known as ultradian rhythms¹⁶, have been observed in rodents with ablated suprachiasmatic nuclei (SCNs)¹⁷, the mouse *Clock* mutant¹⁸ and *Drosophila* clock mutants devoid of circadian rhythms¹⁹. In contrast, the circadian peak predominated in the wild-type animals and in the mutant animals when they displayed a clear circadian rhythm (Fig. 3a, b and data not shown). Thus the *mPer2*^{Brdm1} mutation provides another example where ultradian rhythmicity can be dissociated from circadian rhythmicity.

The loss of circadian rhythmicity in m/m mice is not due to a decrease in total activity. Total wheel-running activity was not significantly different among the three genotypes in constant darkness (for the first and second interval in DD, wheel rotations per day are $18,849 \pm 1,868$ and $21,391 \pm 2,366$, respectively, for +/+ mice,

$n = 13$; $19,833 \pm 2,283$ and $21,108 \pm 2,669$, respectively, for $+/+$ mice, $n = 13$; and $20,899 \pm 2,402$ and $21,708 \pm 2,275$, respectively, for m/m mice, $n = 17$; numbers are mean $\pm$ s.e.m.). The light pulse temporarily restored a circadian rhythm in all m/m animals, which then lasted for a variable period of time (Fig. 2b, c), indicating that the loss of circadian rhythmicity is reversible. Histological analysis of SCN sections showed no gross anatomical difference between $+/+$ and m/m F_2 littermates (data not shown), indicating that the circadian phenotype in the mutants is not due to a developmental defect in the SCN. To determine whether genetic background affects the expressivity of the circadian phenotype, we generated $mPer2^{Brdm1}$ mutants in the 129-inbred strain by crossing chimaeric mice to strain 129 females. The same circadian phenotype was observed in 129-homozygous mutants, that is, a short period for a variable time followed by a loss of circadian rhythmicity (circadian period: $+/+$, 23.6 ± 0.3 h (mean $\pm$ s.d.), $n = 10$; m/m , 22.1 ± 0.5 h, $n = 8$) (for representative activity records, see Supplementary Information, Fig. A). The $mPer2^{Brdm1}$ mutation has disrupted two basic properties of the circadian clock, period length and persistence of circadian rhythmicity, showing that $mPer2$ functions in the mammalian circadian clock.

Although homozygous $mPer2^{Brdm1}$ mutants entrained to the LD cycle, they showed increased locomotor activity in a several-hour period preceding the light-to-dark transition (pre-dusk activity) (Fig. 2c). On average, for F_2 homozygous mutants, $12 \pm 1.4\%$ ($n = 17$, mean $\pm$ s.e.m.) of the total activity was between ZT 9 and ZT 12. (ZT, zeitgeber time. In an LD 12:12 cycle, light is turned on at ZT 0 and off at ZT 12.) In contrast, wild-type littermates displayed just $2 \pm 1.0\%$ ($n = 13$) of their total activity during the same period. Such elevated pre-dusk activity was also observed in 129-homozygous mutants (Supplementary Information, Fig. A). Elevated pre-dusk activity essentially reflects a phase advance in daily activity, which is an expected feature associated with a short period^{20,21}.

To determine whether the mutation also affects the expression of genes that are believed to be components of the circadian system, we examined the expression of $mPer1$, $mPer2$, $mPer3$ and $Clock$ in the SCNs of $+/+$ and m/m littermates, at four different times in the LD cycle, by *in situ* hybridization. Expression of $mPer2$ in wild-type controls peaks at around ZT 12 (Fig. 4a), as previously observed⁴. In contrast, $mPer2$ transcript levels are much lower at ZT 12 in m/m animals (about 20–40% of wild-type levels by silver grain counting). $mPer2$ expression at ZT 6, 18 and 24 is at background levels in the mutants, similar to the wild-type controls. Examination of $mPer2$ expression every two hours between ZT 6 and ZT 14 indicated that an overall decrease in expression levels is accompanied by a slight phase advance of transcript oscillation in the mutants (with the new peak of expression at ZT 10; see Supplementary Information, Fig. B). The slight phase advance in $mPer2$ expression is consistent with a slight phase advance in the daily rhythm. A similar relationship between behavioural and molecular alterations has been documented in *Drosophila per^S* mutants^{20,22}. The concordance of the phase shift in circadian activity and transcript oscillation is an expected feature for an oscillator component that describes the progression of the clock²³. The fact that $mPer2$ mutant transcripts still oscillate in the $mPer2^{Brdm1}$ mutant SCN in LD indicates that $mPer2$ is not solely responsible for its own oscillating expression.

Expression of $mPer1$ in the SCN of wild-type littermates is strong at ZT 6 and declines to background levels at ZT 12, 18 and 24, as previously observed². In contrast, $mPer1$ expression in the mutants is significantly reduced at ZT 6 (10–30% of wild-type control) (Fig. 4a). At ZT 12, 18 and 24, $mPer1$ transcripts are at background levels in the mutants, similar to the wild-type controls. In heterozygous mutants, the expression of $mPer1$ or $mPer2$ is not detectably different from that in the wild-type controls (Supplementary Information, Fig. C, and data not shown). The constitutive expression of $mPer1$ and $mPer2$ in the hippocampus and piriform cortex^{2,4}

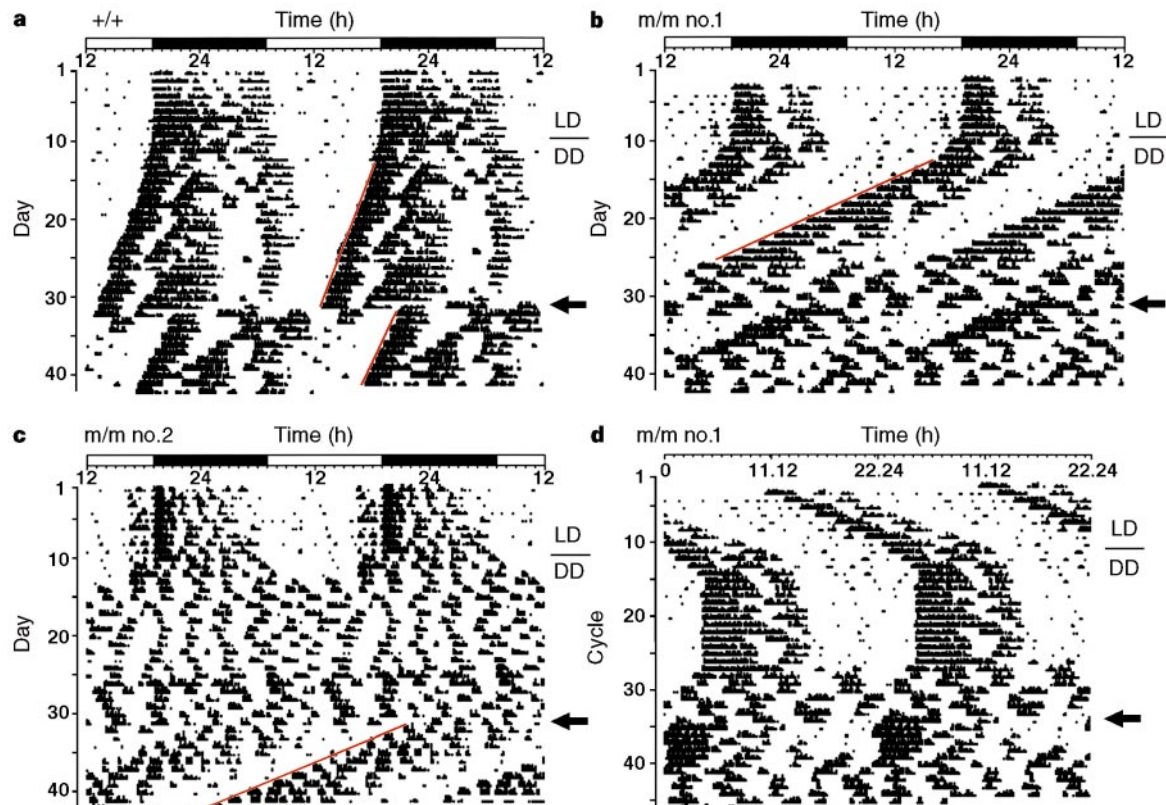


Figure 2 Representative locomotor activity records of F_2 wild-type and homozygous $mPer2^{Brdm1}$ mutant mice. Activity records of a wild-type mouse (a) and two homozygous mutants (b, c) are shown. The bar over the records indicates the LD cycle. The line over DD indicates the transition from LD to DD. A

6-h light pulse was applied between 6 pm and midnight (arrow). The slope of the line (in red) aligned with the points of the onset of activity on a 24-h scale plot reflects the period length. d, The same activity record as in (b) but plotted on a scale corresponding to the period length of this mutant animal.

was not detectably different between $+/+$ and m/m littermates (Supplementary Information, Fig. D), indicating that the alteration of *mPer1* and *mPer2* expression in the mutant SCN is specific to the rhythmic control of the two genes. In peripheral tissues such as liver and kidney, where *mPer* transcript levels also oscillate strongly^{7,24}, there was a similar substantial reduction in the amplitude of *mPer1* and *mPer2* transcript oscillation in homozygous *mPer2^{Brdm1}* mutants (Fig. 4b and data not shown). Examination of *mPer1* and *mPer2* expression in the SCN in constant darkness (as assayed every three hours for a 24 h period in the second day in DD) showed a similar reduction of the oscillating expression of the two genes in the mutant as in LD (data not shown).

The *mPer2^{Brdm1}* mutation greatly reduces the rhythmic expression of *mPer1* and *mPer2*, indicating that *mPer2* regulates *mPer1* *in vivo*. This implies that *mPer2* is upstream of *mPer1*. It is equally possible that the two genes auto- and cross-regulate in a single molecular cycle for circadian regulation. The reduced expression of *mPer1* and *mPer2* in the mutants, rather than elevated expression, is unexpected given the proposed role of the mPers as negative elements in a putative molecular autoregulatory loop²⁵. Instead, our data indicate that *mPer2* has positive regulatory functions in the clock, reminiscent of a recent study where *Drosophila* (d)Per was implicated in positive regulation of *dClock*²⁶. In contrast to the altered *mPer1* and *mPer2* expression, the expression of both *mPer3* and *Clock* was indistinguishable in the mutant SCN from that of the wild-type controls, and did not appear to oscillate at the four time points sampled in both genotypes (Fig. 4a). Therefore, the alteration of *mPer1* or *mPer2* expression by the *mPer2^{Brdm1}* mutation does not appear to work through a change in either *Clock* or *mPer3* expression.

The phenotype of our deletion mutation underscores the func-

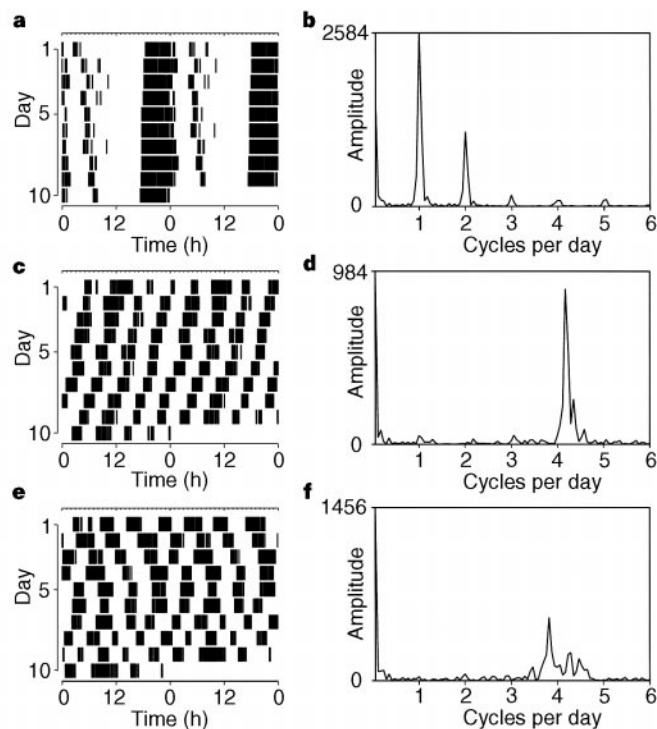


Figure 3 Fourier analysis of periodicity of F_2 wild-type and homozygous *mPer2^{Brdm1}* mutants. Activity records of one wild-type (**a**) and two m/m mice (**c**, **e**) for a 10-day interval in DD, the latter two with no apparent circadian rhythmicity. **b**, **d**, **f**, Fourier periodogram analyses on the activity data plotted in **a**, **c** and **e**, respectively. The amplitude of any rhythm (y-axis) is plotted against the number of cycles per day (x-axis) with long periods (fewer cycles per day) to the left and short periods (more cycles per day) to the right. A peak at around 1 cycle per day indicates the presence of a circadian rhythm (as in **b**).

tional importance of the PAS domain of *mPer2*. Given the significant down-regulation of *mPer1* in the SCN of homozygous mutants, the apparently normal *mPer1* expression in the SCN of heterozygous mutants together with a recessive circadian phenotype is consistent with *mPer2^{Brdm1}* being a loss-of-function mutation. However, it is possible that the *mPer2^{Brdm1}* allele has some residual function.

In conclusion, our data show that *mPer2* encodes a dedicated component of the mammalian circadian clock. Given its cyclic expression in the SCN⁴⁻⁶, our data implicate *mPer2* as an oscillator component of the clock. *mPer2* is probably a central clock component, as mutation in *mPer2* leads to a change in the expression of other genes believed to be associated with the clock (*mPer1*). The circadian rhythm in *mPer2^{Brdm1}* mutants is not lost instantly upon release into constant darkness. Rather, the clock continues to operate for a variable period of time and then collapses. A similar feature has been observed in the mouse *Clock* mutant¹⁸. This characteristic may imply some partial redundancy in the maintenance

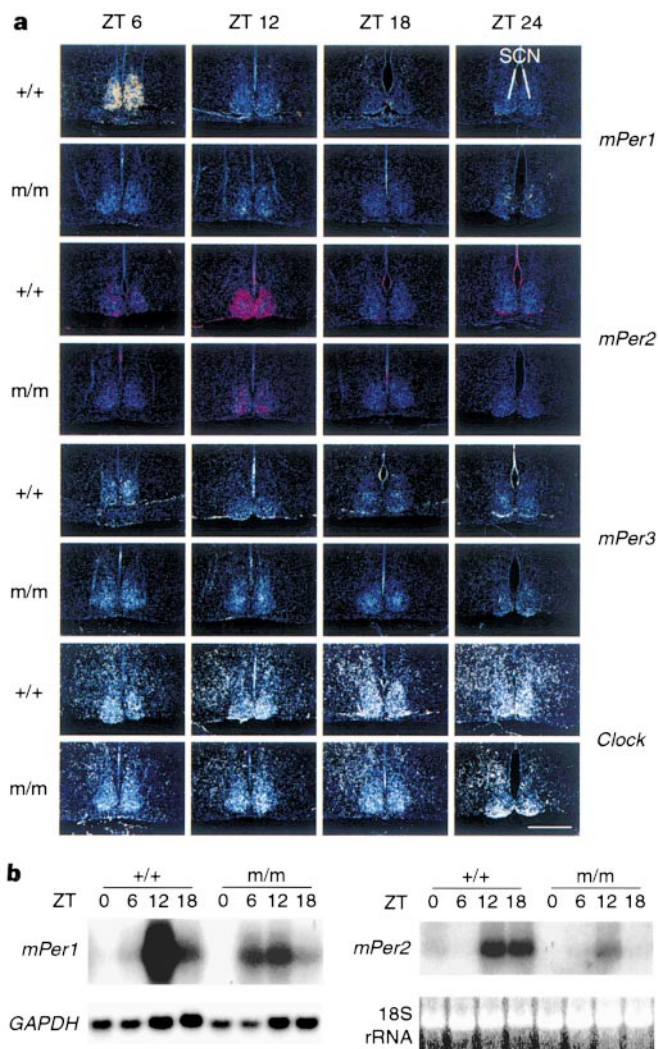


Figure 4 Expression analysis of F_2 wild-type and homozygous *mPer2^{Brdm1}* mutant mice in LD 12:12. **a**, *In situ* hybridization results on SCN of wild-type ($+/+$) and homozygous mutant (m/m) mice at four different time points. The probes are indicated on the right. Yellow, *mPer1*; red, *mPer2*; white, *mPer3* or *Clock*. Scale bar, 500 μ m. **b**, Northern blot analyses of *mPer1* and *mPer2* in liver tissues in $+/+$ and m/m mice. Duplicate blots were hybridized with a full-length *mPer1* cDNA and a full-length *mPer2* cDNA probe respectively. Re-probing of the same blot with a *GAPDH* probe or ethidium bromide staining of 18S ribosomal RNA was used to control for loading and RNA integrity. Representative results are shown from one out of three independent experiments with similar results, each performed on a different set of animals.

of the circadian cycle in mammals. Mutational analyses of other putative clock genes will be essential for unravelling the molecular mechanisms underlying the mammalian circadian clock. These mutants will also provide useful animal models for elucidating the aetiology of and developing treatments for disorders in humans related to the sleep–wake cycle. □

Methods

Generation of *mPer2^{Brdm1}* mutant mice. We isolated a genomic clone from a mouse 129S5/SvEvBrd genomic library using a mouse *mPer2* complementary DNA probe. A targeting vector was constructed with *PGK-Neo* as the positive selection marker and *HSVtk* as the negative selection marker to delete a 2.1-kilobase (kb) fragment. We used a 6.7-kb *Bgl*II fragment as the 5' homology region and a 4.0-kb *Kpn*I fragment as the 3' homology region. The *HSVtk*, *PGK-Neo* and vector backbone were from pKO SelectTK, pKO SelectNeo V800 and pKO Scrambler V924 (Lexicon Genetics). Tissue culture, electroporation, mini-Southern blot analysis on embryonic stem cell colonies, generation of chimaeric and germline mice and tail DNA genotyping were done as described^{27,28}.

Locomotor activity monitoring and circadian phenotype analysis. Mice were housed in individual cages equipped with a running wheel in ventilated, light-tight chambers with controlled lighting. Wheel-running activity was monitored by an on-line PC using the Chronobiology Kit (Stanford Software Systems). In the LD cycle, the light was turned on at 7:00 (ZT 0) and off at 19:00 (ZT 12). The switch into constant darkness (DD) was effected by not turning on the light at the usual time. The activity records are double plotted so that each day/cycle's activity is shown both to the right and below that of the previous day/cycle. Activity is plotted in density percentile distribution (Fig. 2) or threshold (Fig. 3) format. For activity counting we used the ACTCNT program of the Chronobiology Kit. To determine the period length, an interval with a 10-day minimum during which the circadian period appeared to be stable on the activity record was analysed with a χ^2 periodogram²⁹ using the Stanford Chronobiology Kit. We used Fourier periodogram analysis¹⁵ in the Chronobiology Kit to assess the strength of circadian and/or ultradian rhythmicity.

In situ hybridization. Mice were killed by cervical dislocation under ambient light conditions at ZT 6 and ZT 12 and under a 15 W safety red light at ZT 18 and ZT 0/24. Specimen preparation and *in situ* hybridization with an *mPer1* and an *mPer2* probe were carried out as described^{4,30}. The *mPer2* probe is outside the region deleted in the mutant. The *mPer3* probe was made from an RT–PCR product corresponding to nucleotides 480–824 (AF050182). The *Clock* probe was made from an RT–PCR product corresponding to nucleotides 1352–2080 (AF000998). Tissue was visualized by fluorescence of Hoechst dye-stained nuclei (blue in Fig. 4).

Northern blot and RT-PCR analysis. Tissues were collected and frozen in liquid nitrogen and stored at –80 °C. RNA was prepared with the RNAzol™ B RNA isolation kit (TEL-TEST). We performed northern blot analysis on total tissue RNA using denaturing formaldehyde gel. For RT–PCR analysis, first strand cDNA was generated using Moloney reverse transcriptase (BRL-GIBCO) and oligo dT-priming from total liver RNA. An aliquot of the first strand cDNA was then amplified by PCR across the deletion region with the 5' primer CTA CCT GGT CAA GGT GCA AGA G and the 3' primer GGT TTG AAT CTT GCC ACT GG. The RT–PCR products were then sequenced with an internal primer AGG GTA CAC TCG GGC TAT GA.

Received 31 December 1998; accepted 18 May 1999.

- Pittendrigh, C. S. Temporal organization: reflections of a Darwinian clock-watcher. *Annu. Rev. Physiol.* **55**, 16–54 (1993).
- Sun, Z. S. *et al.* *RIGUL*, a putative mammalian ortholog of the *Drosophila period* gene. *Cell* **90**, 1003–1011 (1997).
- Tei, H. *et al.* Circadian oscillation of a mammalian homologue of the *Drosophila period* gene. *Nature* **389**, 512–516 (1997).
- Albrecht, U., Sun, Z. S., Eichele, G. & Lee, C. C. A differential response of two putative mammalian circadian regulators, *mper1* and *mper2*, to light. *Cell* **91**, 1055–1064 (1997).
- Shearman, L. P., Zylka, M. J., Weaver, D. R., Kolakowski, L. F. Jr & Reppert, S. M. Two *period* homologs: circadian expression and photic regulation in the suprachiasmatic nuclei. *Neuron* **19**, 1261–1269 (1997).
- Takumi, T. *et al.* A new mammalian *period* gene predominantly expressed in the suprachiasmatic nucleus. *Genes Cells* **3**, 167–176 (1998).
- Zylka, M. J., Shearman, L. P., Weaver, D. R. & Reppert, S. M. Three *period* homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. *Neuron* **20**, 1103–1110 (1998).

- Takumi, T. *et al.* A light-independent oscillatory gene *mPer3* in mouse SCN and OVLT. *EMBO J.* **17**, 4753–4759 (1998).
- Huang, Z. J., Edery, I. & Rosbash, M. PAS is a dimerization domain common to *Drosophila* *Period* and several transcription factors. *Nature* **364**, 259–262 (1993).
- Crosthwaite, S. K., Dunlap, J. C. & Loros, J. J. *Neurospora wc-1* and *wc-2*: transcription, photo-responses, and the origins of circadian rhythmicity. *Science* **276**, 763–769 (1997).
- King, D. P. *et al.* Positional cloning of the mouse circadian clock gene. *Cell* **89**, 641–653 (1997).
- Rutla, J. E. *et al.* CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila period* and *timeless*. *Cell* **93**, 805–814 (1998).
- Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M. A mutant *Drosophila* homolog of mammalian *Clock* disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell* **93**, 791–804 (1998).
- Ponting, C. P. & Aravind, L. PAS: a multifunctional domain family comes to light. *Curr. Biol.* **7**, R674–R677 (1997).
- Bracewell, R. N. *The Hartley Transform* (Oxford Univ. Press, New York, 1986).
- Aschoff, J. in *Handbook of Behavioral Neurobiology 4: Biological Rhythms* (ed. Aschoff, J.) 3–10 (Plenum, New York, 1981).
- Ibuka, N., Inouye, S. I. & Kawamura, H. Analysis of sleep–wakefulness rhythms in male rats after suprachiasmatic nucleus lesions and ocular enucleation. *Brain Res.* **122**, 33–47 (1977).
- Vitaterna, M. H. *et al.* Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science* **264**, 719–725 (1994).
- Dowse, H. B. & Ringo, J. M. Further evidence that the circadian clock in *Drosophila* is a population of coupled ultradian oscillators. *J. Biol. Rhythms* **2**, 65–76 (1987).
- Hamblen-Coyle, M. J., Wheeler, D. A., Rutla, J. E., Rosbash, M. & Hall, J. C. Behavior of period-altered rhythm mutants of *Drosophila* in light:dark cycles. *J. Insect Behav.* **5**, 417–446 (1992).
- Ralph, M. R. & Menaker, M. A mutation of the circadian system in golden hamsters. *Science* **241**, 1225–1227 (1988).
- Hardin, P. E., Hall, J. C. & Rosbash, M. Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536–540 (1990).
- Dunlap, J. C. Genetics and molecular analysis of circadian rhythms. *Annu. Rev. Genet.* **30**, 579–601 (1996).
- Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).
- Sangoram, A. M. *et al.* Mammalian circadian autoregulatory loop: A *Timeless* ortholog and *mPer1* interact and negatively regulate CLOCK-BMAL1-induced transcription. *Neuron* **21**, 1101–1113 (1998).
- Bae, K., Lee, C., Sidote, D., Chuang, K. Y. & Edery, I. Circadian regulation of a *Drosophila* homolog of the mammalian *Clock* gene: PER and TIM function as positive regulators. *Mol. Cell. Biol.* **18**, 6142–6151 (1998).
- Matzuk, M. M., Finegold, M. J., Su, J. G., Hsueh, A. J. & Bradley, A. Alpha-inhibin is a tumour-suppressor gene with gonadal specificity in mice. *Nature* **360**, 313–319 (1992).
- Ramirez-Solis, R., Davis, A. C. & Bradley, A. Gene targeting in embryonic stem cells. *Methods Enzymol.* **225**, 855–878 (1993).
- Sokolove, P. G. & Bushnell, W. N. The chi square periodogram: its utility for analysis of circadian rhythms. *J. Theor. Biol.* **72**, 131–160 (1978).
- Albrecht, U., Eichele, G., Helms, J. A. & Lu, H. in *Molecular and Cellular Methods in Developmental Toxicology* (ed. Daston, G. P.) 23–48 (CRC Press, Boca Raton, FL, 1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank S. Vaishnav, L. Qiu, Y.-C. Cheah, E. Sheppard and S. Rivera for technical assistance; P. Hastings for comments on the manuscript; J. W. Patrick for providing space and facilities for circadian phenotype analysis; and J. Takahashi, Y. Zhao and M. Bucan for helpful discussions. This work was supported by grants from NINDS and NIDA to J. W. Patrick, from the Max-Planck Society to G. E., from NIH and the Department of Defense to C.C.L., and from NIH and the Howard Hughes Medical Institute to A.B. A.B. is an investigator with HHMI.

Correspondence and requests for materials should be addressed to C.C.L. (e-mail: ching@bcm.tmc.edu)

Immunization with amyloid- β attenuates Alzheimer-disease-like pathology in the PDAPP mouse

Dale Schenk, Robin Barbour, Whitney Dunn, Grace Gordon, Henry Grajeda, Teresa Guido, Kang Hu, Jiping Huang, Kelly Johnson-Wood, Karen Khan, Dora Kholodenko, Mike Lee, Zhenmei Liao, Ivan Lieberburg, Ruth Motter, Linda Mutter, Ferdie Soriano, George Shopp, Nicki Vasquez, Christopher Vandever, Shannan Walker, Mark Wogulis, Ted Yednock, Dora Games & Peter Seubert

Elan Pharmaceuticals, 800 Gateway Boulevard, South San Francisco, California 94080, USA

Amyloid- β peptide (A β) seems to have a central role in the neuropathology of Alzheimer's disease (AD)¹. Familial forms of the disease have been linked to mutations in the amyloid precursor protein (APP) and the presenilin genes^{2,3}. Disease-linked mutations in these genes result in increased production of the

42-amino-acid form of the peptide ($A\beta_{42}$)⁴⁻⁸, which is the predominant form found in the amyloid plaques of Alzheimer's disease^{9,10}. The PDAPP transgenic mouse, which overexpresses mutant human APP (in which the amino acid at position 717 is phenylalanine instead of the normal valine), progressively develops many of the neuropathological hallmarks of Alzheimer's disease in an age- and brain-region-dependent manner^{11,12}. In the present study, transgenic animals were immunized with $A\beta_{42}$, either before the onset of AD-type neuropathologies (at 6 weeks of age) or at an older age (11 months), when amyloid- β deposition and several of the subsequent neuropathological changes were well established. We report that immunization of the young animals essentially prevented the development of β -amyloid-plaque formation, neuritic dystrophy and astrogliosis. Treatment of the older animals also markedly reduced the extent and progression of these AD-like neuropathologies. Our results raise the possibility that immunization with amyloid- β may be effective in preventing and treating Alzheimer's disease.

Our initial experiments investigated the effects of immunization against amyloid-plaque-related proteins on the development of AD-like neuropathology in young PDAPP mice, in which treatment had begun before the occurrence of significant plaque pathology (6 weeks of age). The immunogens were either synthetic human A β ₄₂, the major component of β -amyloid plaques in AD and PDAPP mice, or peptides derived from the primary amino-acid sequence of serum amyloid-P component (SAP). SAP is a protein associated with amyloid plaques in AD and other amyloid diseases¹³. Two additional groups of PDAPP mice were immunized with PBS buffer or were left untreated to serve as controls. Mice received 11 immunizations over an 11-month period. We found that eight of nine PDAPP mice immunized with A β ₄₂ developed and maintained serum antibody titres against A β ₄₂ of greater than 1:10,000. The ninth mouse had a lower titre, of approximately 1:1,000. Cross-reactivity of the antisera was also observed against mouse amyloid- β peptide, although titres were generally an order of magnitude lower. Mice immunized with SAP peptides had antibody titres against SAP ranging from 1:1,000 to 1:10,000, with the exception of one mouse which had a titre greater than 1:10,000 (data not shown). Control animals receiving adjuvant alone had negligible titres to A β ₄₂. At 13

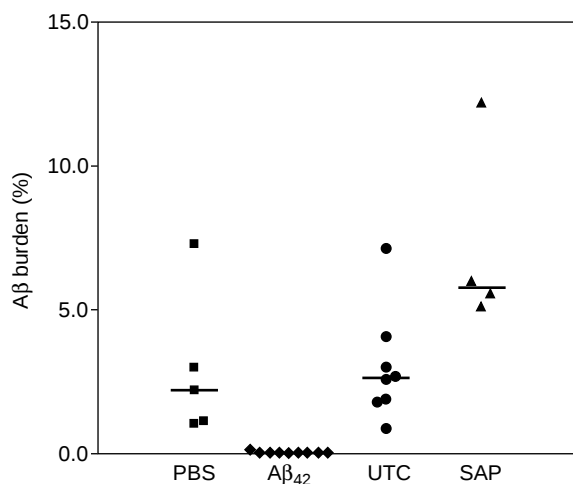


Figure 1 Reduction of A β burden in the hippocampus at 13 months of age in mice immunized with A β ₄₂. PDAPP mice were immunized beginning at 6 weeks of age as described in Methods. The percentage of the area of the hippocampal region occupied by A β deposits was determined by quantitative image analysis. Values for individual mice are shown sorted by treatment group. Horizontal lines represent the median values. The A β ₄₂-immunized group had significantly fewer A β deposits than any of the other three groups ($P = 0.001$), which are not significantly different from each other ($P > 0.05$). UTC, untreated controls; SAP, mice immunized with serum amyloid P.

months of age, quantitative immunohistochemical measures determined the extent of amyloid- β burden and the prevalence of neuritic plaques, astrogliosis and microgliosis.

Immunization with A β ₄₂ resulted in almost complete prevention of amyloid- β deposition (Figs 1, 2). Seven of nine mice immunized with A β ₄₂, including the mouse with the lowest anti-A β titre, had no detectable amyloid- β deposits in their brains. One mouse from this treatment group had a single isolated plaque in the six brain sections examined, whereas a second animal had a greatly reduced amyloid- β burden. Quantitative imaging of the amyloid- β burden in the hippocampus verified the near-total reduction achieved in A β ₄₂-treated animals (Fig. 1). The median values of the amyloid- β burden for the PBS group (2.22%) and for the untreated control group (2.65%) were significantly greater than that of the A β -immunized group (0.00%, $P = 0.0005$). In contrast, the median value of amyloid- β for the group immunized with SAP peptides (5.74%), although increased, was not significantly different from that of control animals. Brain tissue from the PBS-treated control mice contained numerous amyloid- β deposits in the hippocampus (Fig. 2) and retrosplenial cortex (not shown). We observed a similar pattern of amyloid- β deposition in mice immunized with SAP peptides. Brain sections from A β ₄₂-immunized mice were also stained with thioflavin S to rule out the possibility that the lack of immunohistochemically detectable plaques was due to competition by the *de novo* anti-A β antibodies produced by these animals (see below). Thioflavin S detected no amyloid- β deposits in these A β ₄₂-immunized mice (not shown). In contrast, immunization with SAP peptides did not affect amyloid- β deposition, suggesting that

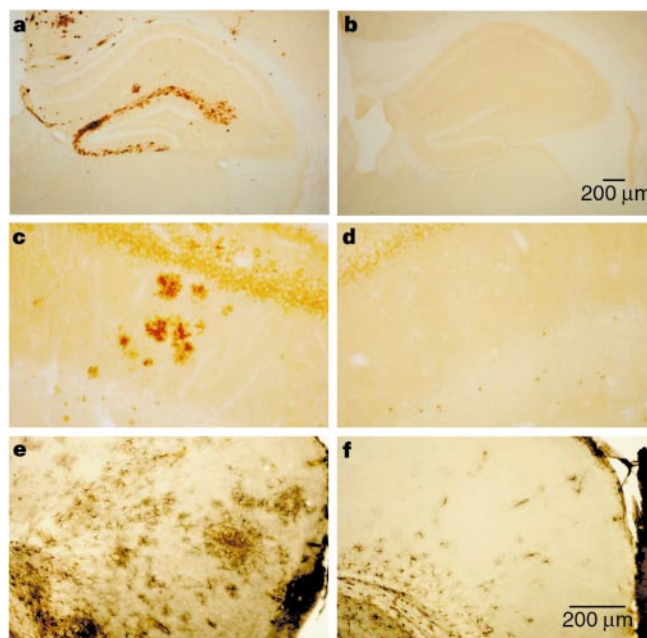


Figure 2 Hippocampal A β deposition, neuritic plaque formation and cortical astrocytosis in PBS- and A β_{42} -injected mice. Images from 13-month-old mice with A β burdens representative of the median values (see Fig. 1) of their respective groups are shown. **a, b**, Hippocampal A β plaques in PBS- (**a**) and A β_{42} -injected (**b**) mice. **a** shows abundant A β deposition in the outer molecular layer of the hippocampal dentate gyrus of a PBS-treated animal. **b** shows no detectable A β in this region of an A β_{42} -immunized mouse, a profile observed in most animals from this group. Scale bar in **b** corresponds to both **a** and **b**. **c, d**, Hippocampal sections from PBS- (**c**) and A β_{42} -injected (**d**) mice. Typical dystrophic neurites associated with neuritic plaques and labelled with the APP-specific monoclonal antibody 8E5 (ref. 16) were found in PBS- (**c**), but not A β_{42} -injected animals (**d**). **e, f**, Abundant plaque-associated astrocytosis, as determined by GFAP immunohistochemistry (Sigma), was evident in the retrosplenial cortex of PBS- (**e**) but not A β_{42} -injected (**f**) mice. Scale bar in **f** corresponds to **c-f**.

immune responses against plaque components *per se* are insufficient to prevent or eliminate β -amyloid plaques and neuropathology.

The brains of $A\beta_{42}$ -treated mice that contained no amyloid- β deposits were also devoid of the dystrophic neurites that characterize the neuritic plaques (Fig. 2: compare c and d). Small numbers of dystrophic neurites were present in the two $A\beta_{42}$ -treated mice that had detectable $A\beta$ deposits. In contrast, all brains from SAP-injected mice and the two control groups (PBS-injected and untreated mice) had numerous neuritic plaques. Image analyses of the hippocampus demonstrated the virtual elimination of dystrophic neurites in the $A\beta_{42}$ -treated mice (median, 0.00%) compared with the PBS recipients (median, 0.28%; $P = 0.0005$).

Astrocytosis, another hallmark of plaque-associated pathology in both Alzheimer's disease and PDAPP mice, was dramatically reduced in the brains of all of the $A\beta_{42}$ -injected mice (Fig. 2). Brains from mice in all other groups contained numerous clusters of astrocytes that were immunoreactive to glial fibrillary acidic protein (GFAP), a finding typical of $A\beta$ -plaque-associated gliosis (Fig. 2e).

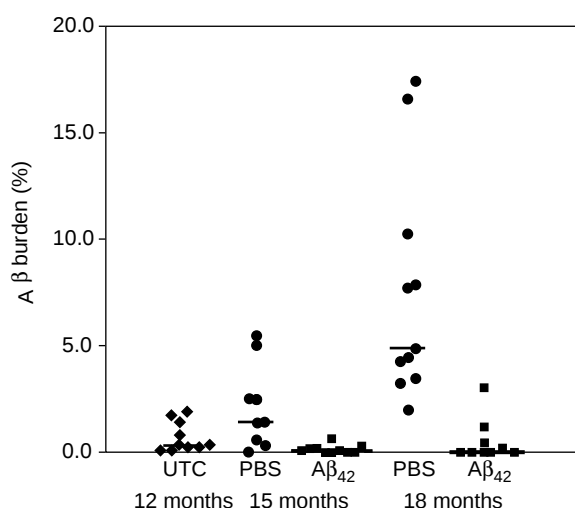


Figure 3 Quantitative image analysis of the cortical $A\beta$ burden in older PBS- and $A\beta$ -treated mice. Immunization of PDAPP mice was begun at 11 months of age. Amyloid burden was significantly reduced in the $A\beta_{42}$ group compared with the PBS controls at both 15 ($P = 0.003$) and 18 ($P = 0.0002$) months of age. The median value of the amyloid burden for each group is shown by the horizontal lines.

Association of amyloid plaques and reactive astrocytes was verified in a subset of GFAP-reacted sections counterstained with thioflavin S. The results of image analyses for the retrosplenial cortex verified that the reduction in astrocytosis was significant, with a median value of 1.55% for mice immunized with $A\beta_{42}$ compared with median values of greater than 6% for groups immunized with SAP peptide or PBS, or untreated control mice ($P = 0.0017$).

Sections of the mouse brains were also reacted with a monoclonal antibody specific for MAC-1 (CD11b; Chemicon), a cell-surface marker that is upregulated on activated, plaque-associated microglia. MAC-1 labelling was substantially lower in the brains of mice treated with $A\beta_{42}$ compared with the PBS control group, a finding consistent with the lack of a plaque-induced gliosis (not shown).

The almost complete absence of plaques in the brains of $A\beta_{42}$ -treated mice indicates that a fundamental mechanism of amyloid plaque formation has been disrupted. Subsequent studies indicate that $A\beta$ production itself was unaffected (data not shown). $A\beta_{42}$ immunization therefore either prevents deposition and/or enhances the clearance of $A\beta$ from the brain. The absence of neuritic and gliotic changes suggests that the $A\beta_{42}$ -immunized mice never developed the neurodegenerative lesions that typify the progression of AD-like pathology in this model. The absence of enhanced astrocytosis, in particular, suggests that the processes preventing β -amyloidosis do not in themselves cause appreciable damage to the neuropil.

The above results clearly indicate that $A\beta_{42}$ immunization essentially prevents the development of AD-like neuropathology in the PDAPP mouse. It was unclear whether $A\beta_{42}$ immunization would improve the pathological outcome if treatment were initiated when a substantial $A\beta$ plaque burden already existed. We therefore undertook further experiments in which immunizations with $A\beta_{42}$ began at an age when many β -amyloid plaques are already present in the brains of the PDAPP mice. Immunizations were continued during a period when the extent of $A\beta$ deposition reaches levels comparable to those of established AD^{11,12}. For this study, approximately 11-month-old, heterozygotic female PDAPP mice ($n = 24$) were immunized repeatedly with $A\beta_{42}$ plus adjuvant, with an injection protocol and schedule equivalent to those used in the young PDAPP mouse study (see Methods). Similar titre responses against $A\beta_{42}$ to those seen in younger animals were generated in the older animals. As a negative control, a parallel group of 24 transgenic littermates was immunized with PBS plus adjuvant. One-half of each group was killed at 15 months of age after 4 months' treatment, and the remaining half was killed at 18 months

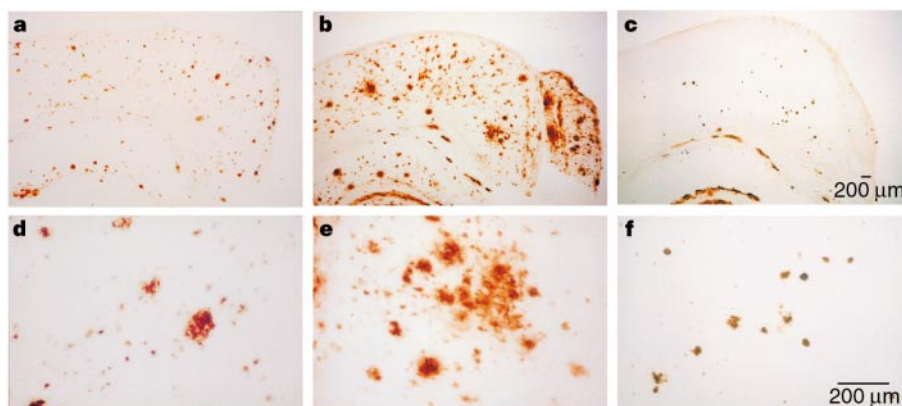


Figure 4 Reduction of cortical $A\beta$ deposition in older PDAPP mice immunized with $A\beta_{42}$. $A\beta$ deposits in the brains of 12-month-old untreated PDAPP mice and 18-month-old PBS- and $A\beta_{42}$ -injected mice with median $A\beta$ burdens representative of their respective treatment groups (see Fig. 3). **a**, Distribution of amyloid plaques in the frontal and retrosplenial cortices of a 12-month-old untreated PDAPP mouse, typifying the plaque burden of both PBS- and $A\beta_{42}$ -injected groups at

the start of the study. The plaques are shown at a high magnification in **d**. Compared with the 18-month PBS controls (**b**, **e**), $A\beta$ deposits were significantly decreased in the 18-month $A\beta_{42}$ -immunized group (**c**, **f**). Most of the cortical $A\beta$ in brains of $A\beta_{42}$ -injected mice was detected in small extracellular or cell-associated deposits (**f**) compared with the large and numerous extracellular deposits in the PBS group (**e**). Scale bar in **c** corresponds to **a-c**. Scale bar in **f** corresponds to **d-f**.

of age after 7 months' treatment. Groups of untreated PDAPP mice were also killed at ages 12, 15 and 18 months to serve as age-matched, non-immunized controls. At each time point, brains were examined by image analysis and enzyme-linked immunosorbent assay (ELISA) to determine the magnitude of the amyloid- β burden and the extent of neuritic dystrophy, astrogliosis and microgliosis.

Figure 3 shows the results of $A\beta_{42}$ treatment on cortical amyloid- β burden, determined by quantitative image analysis. The median value of cortical amyloid- β burden was 0.28% in untreated, 12-month-old PDAPP mice, a value representative of the plaque load in the experimental mice at the start of the study. At 18 months, the amyloid- β burden had increased by more than 17-fold to 4.87% in PBS-treated mice, while $A\beta_{42}$ -treated mice had a greatly reduced amyloid burden of only 0.01%. The amyloid- β burden was significantly reduced in the $A\beta_{42}$ recipients at both 15 months (96% reduction; $P = 0.003$) and 18 months (>99% reduction; $P = 0.0002$). Figure 4 depicts the burden of amyloid- β in the 12-month-old PDAPP brain at the start of the experiment (Fig. 4a, d). At 18 months of age, the progression of $A\beta$ -plaque pathology was obvious in the PBS group (Fig. 4b, e), but greatly diminished in the $A\beta_{42}$ -injected mice (Fig. 4c, f). Compared with the 12-month-old untreated mice, several brains from the $A\beta_{42}$ group had fewer diffuse and mature amyloid- β deposits at 15 and 18 months, suggesting that the treatment had resulted in the clearance of pre-existing amyloid- β deposits (Fig. 4: compare a and d with c and f). Immunohistochemistry with anti-mouse immunoglobulin antibodies showed that the remaining plaques were often decorated

with IgG in $A\beta_{42}$ -treated but not in PBS-treated mice at both 15 and 18 months of age (data not shown).

Cortical amyloid- β deposition in PDAPP brains begins in the cingulate, frontal and retrosplenial cortices and progresses in a lateral-ventral fashion to sequentially involve the temporal and entorhinal cortices. After 3 months of treatment, amyloid-plaque pathology was diminished in the retrosplenial cortex (Fig. 5a, c) and completely absent in the entorhinal cortex (Fig. 5b, d) of the $A\beta_{42}$ -injected mice. The progressive β -amyloidosis that would normally pervade the entorhinal cortex was thus halted by $A\beta_{42}$ immunization.

PDAPP mice also invariably develop heavy amyloid- β deposition in the outer molecular layer of the hippocampal dentate gyrus¹¹. In a number of brains from $A\beta_{42}$ -immunized mice, this pattern was considerably altered; the hippocampal deposition no longer contained diffuse amyloid- β deposits, and the banded pattern was completely disrupted (Fig. 6). Instead, unusual punctate structures were present that were reactive with anti- $A\beta$ antibodies, and several appeared to be $A\beta$ -containing cells (Fig. 6b). The pattern of apparent cellular labelling produced by the amyloid- β antibodies was replicated in adjacent sections by immunolabelling with antibodies directed at major histocompatibility complex (MHC) class II molecules (Fig. 6c). Phenotypically, these cells resembled activated microglia and monocytes and were occasionally found associated with the wall or lumen of blood vessels. They were not immunoreactive with antibodies that recognize T-cell (CD43, CD3e) or B-cell (CD45RA, CD45RB) surface markers (data not shown). No such cells were found in any of the PBS-treated mice.

Amyloid- β ELISA analysis of the older PDAPP mice was consistent with the immunohistochemical observations. In untreated PDAPP mice, the median level of total $A\beta$ in the cortex at 12 months was 1,600 ng per g (wet weight); this had increased to 8,700 ng per g by 15 months¹². At 18 months the value was 22,000 ng per g, an increase of more than 10-fold during the course of the experiment¹². PBS-treated animals had comparable levels of total amyloid- β at 15 months (8,600 ng per g), and 19,000 ng per g at 18 months. In contrast, $A\beta_{42}$ -treated animals had 81% less total amyloid- β at 15 months (1,600 ng per g) than the PBS-immunized group. Significantly less total amyloid- β (5,200 ng per g) was found at 18 months when the $A\beta_{42}$ and PBS groups were compared, representing a 72% reduction ($P = 0.0001$) in the amyloid- β that would have otherwise been present. Similar results were obtained when cortical levels of $A\beta_{42}$ were compared, namely that the $A\beta_{42}$ -treated group contained much less $A\beta_{42}$, and the differences between the $A\beta_{42}$ and PBS groups were significant at 15 months ($P = 0.04$) and 18 months ($P = 0.0001$). In contrast, cortical levels of APP decreased by less than 10% of those in 12-month-old untreated animals (data not shown)¹². These findings argue against the possibility that the reduction in amyloid- β deposition seen in the treated mice is due to an alteration in APP metabolism.

The progression of neuritic pathology was significantly reduced in the frontal cortex of $A\beta_{42}$ -treated compared with PBS-treated mice (Table 1). At 15 months of age, the neuritic plaque burden in

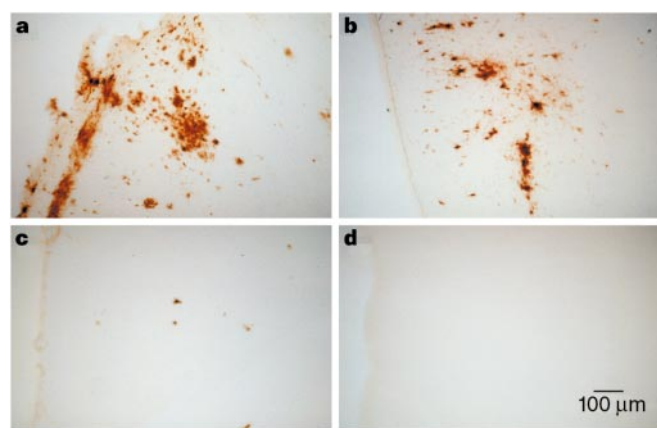


Figure 5 Reduction of $A\beta$ burden in the entorhinal and retrosplenial cortex of older PDAPP mice following $A\beta$ injection. $A\beta$ deposition in the retrosplenial (RSC) and entorhinal (EC) cortices of 15-month-old PBS- (a, b) and $A\beta_{42}$ - (c, d) injected mice with $A\beta$ burdens representative of the median values of their respective groups. $A\beta$ deposition was greatly reduced in the RSC of $A\beta_{42}$ -injected mice compared with the PBS group (compare a and c). No $A\beta$ was detected in the EC of $A\beta_{42}$ -injected mice (d), in contrast to the PBS group (b). Scale bar in d corresponds to all panels.

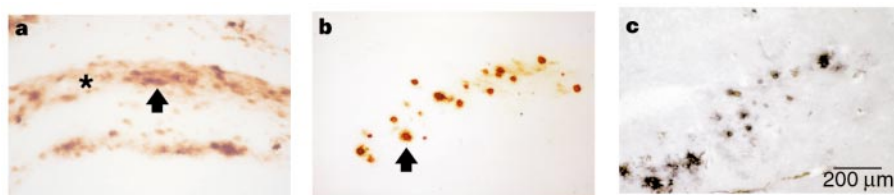


Figure 6 Altered $A\beta$ burden in the hippocampus of older $A\beta_{42}$ -treated mice. Distribution of hippocampal $A\beta$ in $A\beta_{42}$ -injected brains (b), compared with the PBS group (a) at 18 months of age. In the PBS group, the characteristic appearance of diffuse (asterisk) and compacted (arrow) $A\beta$ deposits was evident (a). This pattern was markedly altered in a number of $A\beta_{42}$ -immunized mice (b), with the absence of diffuse deposits and an unusual punctate pattern of $A\beta$ immuno-

reactivity associated with cells (b, arrow). The distribution of MHC II-labelled (Pharmingen) cells in a near-adjacent section (c) corresponded to the pattern of $A\beta$ -positive cellular labelling shown in b. No such obvious cell $A\beta$ staining was found in the PBS group (compare a and b). Scale bar in c corresponds to all panels.

Table 1 Image analysis of neuritic plaque and astrocytic burden in A β ₄₂-treated PDAPP mice

	15 months		18 months	
	PBS	A β ₄₂ -treated	PBS	A β ₄₂ -treated
Neuritic plaque (%)	0.32	0.02	0.49	0.22
Astrocytosis (%)	4.26	1.89	5.21	3.20

Quantitative image analysis of neuritic plaques and astrocytosis was performed using antibody 8E5 to human APP and anti-GFAP, respectively. Methods are described in Fig. 1 legend. Reduction of neuritic plaque burden at ages 15 and 18 months by A β ₄₂ treatment was statistically significant ($P = 0.03$ and 0.01 , respectively), as was reduction of astrocytosis at the same time points ($P = 0.01$ and 0.03 , respectively).

the A β ₄₂-treated mice was reduced by 84% compared with the PBS group (0.05% and 0.32%; $P = 0.03$). Reduction in the neuritic-plaque pathology was similarly maintained between the two groups at 18 months of age, where the degree of neuritic dystrophy was reduced by 55% in the A β ₄₂-treated mice (0.22% and 0.49%; $P = 0.01$).

Reactive astrocytosis was also significantly reduced in the retrosplenial cortex of A β ₄₂-treated mice compared with PBS-treated mice at both 15 and 18 months of age. The per cent of astrocytosis in the PBS group increased between 15 and 18 months from 4.29% to 5.21%. A β ₄₂ treatment suppressed the development of astrocytosis at both time points to 1.89% and 3.2%, respectively. These differences represent a 56% reduction ($P = 0.01$) at 15 months of age and a 34% decrease ($P = 0.03$) at 18 months of age in the A β ₄₂-treated group.

In summary, immunization with A β ₄₂ greatly reduced the development of the AD-like pathology that otherwise occurs in the PDAPP mouse. Immunization preceding plaque development profoundly affected the occurrence of new lesions, as the progression of β -amyloidosis and associated neuropathology was essentially wholly blocked, as seen both in the entire brain of the young animals and in at-risk brain regions of the older animals. Amyloid- β immunization also significantly retarded the progression of existing pathology in affected regions of the older animals. Outcomes of A β -plaque burden, neuritic dystrophy and gliosis were all significantly improved by A β ₄₂ treatment in both young and old animals. In addition, the mechanism resulting in plaque reduction did not seem to produce any obvious signs of damage to the neuropil of A β ₄₂-immunized animals. Histological examination of several organs, including brain and kidney, revealed no signs of immune-mediated complications, despite the high levels of human APP expressed in these tissues and the significant antibody titre to endogenous mouse A β peptide (data not shown).

To our knowledge, this is the first report of a clinically relevant treatment that reduces the progression of AD-like neuropathology in a transgenic animal model of the disease. Although it remains unproven, it is not unreasonable to expect that a similar reduction of neuropathology in AD patients would be of clinical benefit. Although our understanding of the precise aspects of the immune response that result in reduced pathology is incomplete, we have shown that A β ₄₂ immunization results in the generation of anti-A β antibodies and that A β -immunoreactive monocytic/microglial cells appear in the regions of remaining plaques. Thus, one possible mechanism of action is that anti-A β antibodies facilitate clearance of amyloid- β either before deposition, or after plaque formation, by triggering monocytic/microglial cells to clear amyloid- β using signals mediated by Fc receptors.

It has been suggested that a chronic inflammatory state exists in the brain of patients with Alzheimer's disease: specifically the levels of complement, cytokines and acute-phase proteins are raised¹⁴. These observations have led to the hypothesis that anti-inflammatory regimens might be of therapeutic value. The findings presented here argue that an alternative approach, one that augments a highly specific immune response, can markedly reduce pathology in an animal model of the disease. Collectively, the results suggest that

amyloid- β immunization may prove beneficial for both the treatment and prevention of Alzheimer's disease. □

Methods

Immunization procedures. A β peptide was freshly prepared from lyophilized powder for each set of injections. For immunizations, 2 mg A β ₄₂ (human A β ₁₋₄₂; US Peptides) was added to 0.9 ml deionized water and the mixture was vortexed to generate a relatively uniform suspension. A 100- μ l aliquot of 10 $\times$ PBS (where 1 $\times$ PBS is 0.15 M NaCl, 0.01 M sodium phosphate, pH 7.5) was added. The suspension was vortexed again and incubated overnight at 37°C for use the next day. Serum amyloid-P component immunogens were prepared using mouse SAP amino acids 77–85 and 164–173, each conjugated to sheep anti-mouse IgG (Jackson Immunochemicals) as described¹⁵. A β ₄₂ or SAP peptides (100 μ g antigen per injection) were emulsified 1:1 (v/v) with complete Freund's adjuvant for the first immunization, followed by a boost in incomplete Freund's adjuvant at 2 weeks, and monthly thereafter. A β ₄₂ or SAP in PBS alone was injected from the fifth immunization onward. Titres were determined by serial dilutions of sera against either aggregated A β ₄₂ or SAP protein which had been coated onto microtitre wells. Detection used goat anti-mouse immunoglobulin conjugated to horseradish peroxidase and slow-TMB (3,3',5,5'-tetramethyl benzidine; Pierce) substrate. Titres were defined as the dilution yielding 50% of the maximal signal.

Neuropathology quantification. To quantify amyloid burden, PDAPP mouse brain tissue was fixed in 4% paraformaldehyde, cut to 40- μ m coronal sections and reacted with an anti-A β biotinylated antibody, 3D6, as described previously¹¹. Quantitative image analysis was performed using a Videometric 150 Image Analysis System (Oncor) linked to a Nikon Microphot-FX microscope through a CCD video camera. The image of the immunoreacted section was stored in a video buffer and a specific brain region (hippocampus or cortex) was manually outlined and the total pixel area occupied by the structure determined. A monochromatic-based threshold was set to select pixels corresponding to immunolabelled structures. The per cent of the brain region occupied by the labelled pixels was then calculated. For all image analyses, six sections at the level of the dorsal hippocampus, each separated by consecutive 240- μ m intervals, were evaluated for each animal. In all cases, the treatment status of the animals was unknown to the observer. Mann–Whitney nonparametric analysis was performed using Statview software (SAS Institute, Cary, NC). Similar methodologies were used to quantify neuritic dystrophy and gliosis. Specific reagents are indicated in the appropriate figures.

Received 6 April; accepted 5 May 1999.

1. Hyman, B. T. New neuropathological criteria for Alzheimer disease. *Arch. Neurol.* **55**, 1174–1176 (1998).
2. Tanzi, R. E. *et al.* The gene defects responsible for familial Alzheimer's disease. *Neurobiol. Dis.* **3**, 159–168 (1996).
3. Hardy, J. New insights into the genetics of Alzheimer's disease. *Ann. Med.* **28**, 255–258 (1996).
4. Citron, M. *et al.* Mutation of the β -amyloid precursor protein in familial Alzheimer's disease increases β -protein production. *Nature* **360**, 672–674 (1992).
5. Scheuner, D. *et al.* Secreted amyloid β -protein similar to that in the senile plaques of Alzheimer's disease is increased *in vivo* by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nature Med.* **2**, 864–870 (1996).
6. Suzuki, N. *et al.* An increased percentage of long amyloid β protein secreted by familial amyloid β precursor protein (APP717) mutants. *Science* **264**, 1336–1340 (1994).
7. Citron, M. *et al.* Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β -protein in both transfected cells and transgenic mice. *Nature Med.* **3**, 67–72 (1997).
8. Borchelt, D. R. *et al.* Familial Alzheimer's disease-linked presenilin 1 variants elevate A β ₁₋₄₂/A β ₁₋₄₀ ratio *in vitro* and *in vivo*. *Neuron* **17**, 1005–1013 (1996).
9. Iwatsubo, T. *et al.* Visualization of A β ₄₂(43) and A β ₄₀ in senile plaques with end-specific A β monoclonals: evidence that an initially deposited species is A β ₄₂(43). *Neuron* **13**, 45–53 (1994).
10. Lippa, C. F., Nee, L. E., Mori, H. & St George-Hyslop, P. A β ₄₂ deposition precedes other changes in PS-1 Alzheimer's disease. *Lancet* **352**, 1117–1118 (1998).
11. Games, D. *et al.* Alzheimer-type neuropathology in transgenic mice overexpressing V717F β -amyloid precursor protein. *Nature* **373**, 523–527 (1995).
12. Johnson-Wood, K. *et al.* Amyloid precursor protein processing and A β ₄₂ deposition in a transgenic mouse model of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **94**, 1550–1555 (1997).
13. Liang, J. S. *et al.* Evidence for local production of acute phase response apolipoprotein serum amyloid A in Alzheimer's disease brain. *Neurosci. Lett.* **225**, 73–76 (1997).
14. McGeer, E. G. & McGeer, P. L. The role of the immune system in neurodegenerative disorders. *Mov. Disorders* **12**, 855–858 (1997).
15. Seubert, P. *et al.* Isolation and quantification of soluble Alzheimer's β -peptide from biological fluids. *Nature* **359**, 325–327 (1992).
16. Seubert, P. *et al.* Secretion of β -amyloid precursor protein cleaved at the amino terminus of the β -amyloid peptide. *Nature* **361**, 260–263 (1993).

Acknowledgements. We thank Rae Lyn Burke for helpful comments.

Correspondence and requests for materials should be addressed to D.S. (e-mail: dschenk@elanpharma.com).

Generation of GTP-bound Ran by RCC1 is required for chromatin-induced mitotic spindle formation

Rafael E. Carazo-Salas^{*}, Giulia Guarguaglini^{*†}, Oliver J. Gruss, Alexandra Segref, Eric Karsenti & Iain W. Mattaj

EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

^{*} These authors contributed equally to this work

Chromosomes are segregated by two antiparallel arrays of microtubules arranged to form the spindle apparatus. During cell division, the nucleation of cytosolic microtubules is prevented and spindle microtubules nucleate from centrosomes (in mitotic animal cells) or around chromosomes (in plants and some meiotic cells)^{1,2}. The molecular mechanism by which chromosomes induce local microtubule nucleation in the absence of centrosomes is unknown^{3–5}, but it can be studied by adding chromatin beads to *Xenopus* egg extracts⁶. The beads nucleate microtubules that eventually reorganize into a bipolar spindle. RCC1, the guanine-nucleotide-exchange factor for the GTPase protein Ran, is a component of chromatin. Using the chromatin bead assay, we show here that the activity of chromosome-associated RCC1 protein is required for spindle formation. Ran itself, when in the GTP-bound state (Ran-GTP), induces microtubule nucleation and spindle-like structures in M-phase extract. We propose that RCC1 generates a high local concentration of Ran-GTP around chromatin which in turn induces the local nucleation of microtubules.

The GTPase Ran, like other small GTPases, relies on both a GTPase-activating protein (GAP) and a guanine-nucleotide-exchange factor (GEF) for its activity. The only known GAP for Ran, ranGAP, is excluded from the nucleus, whereas RCC1, which is the GEF that is required for GDP/GTP exchange on Ran, is a nuclear, chromatin-bound protein^{7–9}. ranGAP acts together with a second family of cytoplasmic proteins that bind to GTP-bound Ran, for example ranBP1, to induce GTP hydrolysis by Ran^{7–9}. Ran is critical for nucleocytoplasmic transport^{7–9}, but additional functions have been proposed for Ran, and one Ran-binding protein, ranBPM, is a centrosomal component¹⁰. Overexpression of ranBPM induces ectopic microtubule nucleation in the cytoplasm, indicating that Ran could be involved in the spatial regulation of microtubule assembly¹⁰ and thus in the regulation of microtubule nucleation during mitosis.

To test this idea, we added excess Ran to mitotic extracts in the presence of centrosomes (see Methods and figure legends for details of the amounts of proteins used). On addition of RanQ69L loaded with GTP (RanQ69L is a mutant form of Ran that cannot hydrolyse GTP^{11,12}), many asters were generated. Although the appearance of these asters was similar to those in control experiments, there were more asters in extracts containing RanQ69L-GTP than in controls with centrosomes alone (data not shown). We investigated this by repeating the experiment in the absence of added centrosomes and found that there was no microtubule assembly either in control extracts (Fig. 1a) or in extracts to which RanT24N, a mutant form of Ran with a low affinity for both GTP and GDP¹², had been added (Fig. 1b). By contrast, when either RanQ69L-GTP or wild-type Ran loaded with GTP- γ S were added to the extract, abundant micro-

tubule-containing structures that resembled either asters or more complex figures were seen (Fig. 1c, d). On addition of RanQ69L-GTP, an average of 103 microtubule assemblies per μ l extract were seen (3 experiments; range, 60–130). For quantification of Fig. 1 and of subsequent figures, only assemblies resembling asters or spindle poles were counted. For Ran-GTP γ S, the corresponding values were 14.5, 10.2 and 21.8. These structures were not observed when Ran11 GTPase¹³, loaded with either GTP- γ S or GDP, was added to the extracts, or when Ran-GDP was added (data not shown).

These results indicated that Ran-GTP could act as a signal to initiate microtubule assembly in egg extracts. The extracts support nuclear-protein import¹⁴, indicating that endogenous Ran is predominantly in the GDP-bound form¹⁵. This means that ranBP1 and ranGAP, which both stimulate Ran's hydrolytic activity, overpower the action of RCC1, which reloads Ran with GTP. To determine whether endogenous Ran also induces microtubule assembly, we added an excess of recombinant RCC1 to the extracts to convert endogenous Ran into the GTP form. We found that this caused microtubule-containing structures to form, confirming that Ran-GTP induces microtubule assembly in M-phase extracts (Fig. 1e). An average of 142 assemblies per μ l were seen (3 experiments; range, 106–200). Two controls demonstrate the specificity of this effect: first, RanT24N, which binds in its nucleotide-free state to RCC1 and inhibits its exchange activity^{12,16,17}, prevents the induction of microtubule assembly by exogenous RCC1 (Fig. 1f); second, the RCC1 effect is reversed by addition of excess ranBP1 or ranGAP, which converts any Ran-GTP generated by RCC1 back to Ran-GDP (data not shown). These results show that Ran-GTP induces microtubule

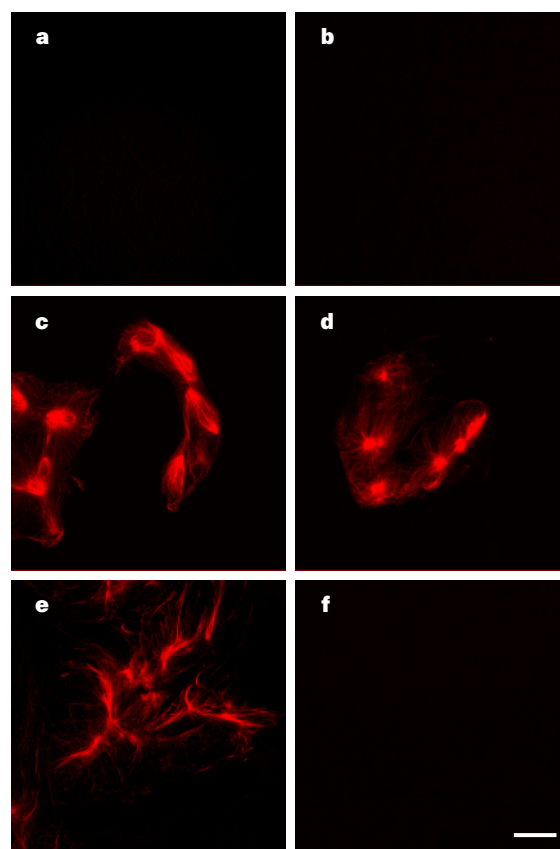


Figure 1 GTP-bound Ran signals microtubule assembly in M-phase extracts. The presence of microtubules was monitored in extracts supplemented with rhodamine-labelled tubulin (red fluorescence). For all panels, M-phase extracts were used after incubation at 20°C for 30 min. **a**, Buffer control; **b**, with 5-fold excess of RanT24N over endogenous Ran; **c**, with 5-fold excess of RanQ69L-GTP; **d**, with 5-fold excess of wild-type Ran loaded with GTP- γ S; **e**, with 20-fold excess of RCC1; **f**, as **e**, but with a 7-fold excess of RanT24N as well. Scale bar, 25 μ m.

[†] Permanent address: CNR, Centre of Evolutionary Genetics, University "La Sapienza", Via degli Apuli 4, 00185 Rome, Italy.

nucleation in M-phase extracts and that RCC1 and ranGAP/ranBP1 specifically regulate the amount of microtubule nucleation by affecting the relative concentrations of Ran-GTP and Ran-GDP in the extracts.

We tested whether this ability of Ran-GTP to include microtubule nucleation in M-phase extracts was related to the capacity of chromosomes or chromatin to induce microtubule nucleation by using a chromatin-bead assay⁶. Microtubule assembly was found to be dependent on a minimum number of clustered beads; in control experiments, no microtubules were seen on 100 clusters of less than seven beads, whereas all 173 clusters of more than 15 beads had associated microtubules (Figs 2b and 3a).

To find out whether RCC1 was among the proteins tightly associated with chromatin beads, we used an anti-RCC1 antibody¹⁸, with anti-ranGAP or anti-ranBP1 antibodies as controls^{19,20} (Fig. 2a) and quantified the signals. Assuming that RCC1 is uniformly distributed, the concentration of RCC1 in the bead fraction was 100-fold more than in the soluble extract. This underestimates the local concentration differential with respect to the extract as both chromatin and RCC1 are confined to the bead surface. RCC1 exchange activity was readily detectable using the bead fraction in *in vitro* exchange reactions (data not shown).

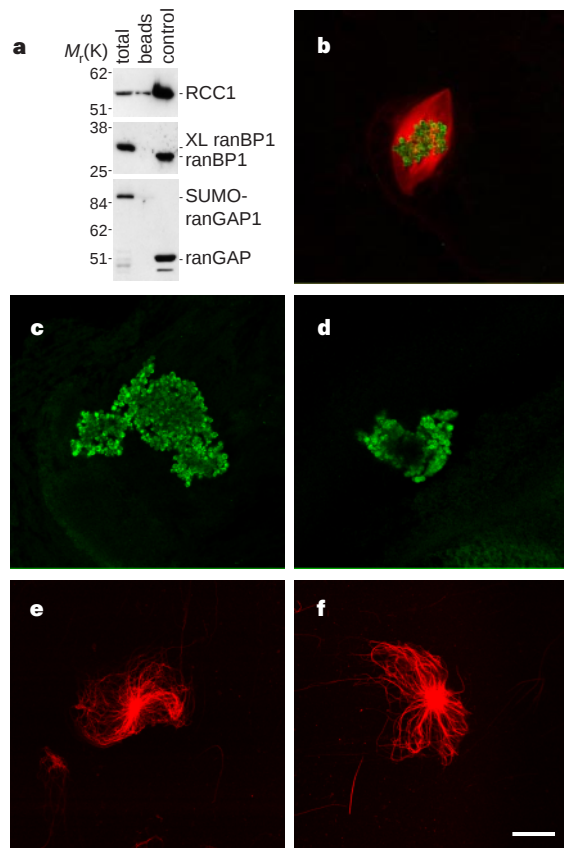


Figure 2 Local RCC1 activity is required for chromatin-induced microtubule assembly. **a**, Western blot of proteins bound to chromatin beads. Recombinant RCC1, ranBP1 and ranGAP were used as controls: the left lane shows these three proteins in 2 μ l of extract (total). Relative molecular mass markers are indicated on the left. Owing to SUMO modification of ranGAP in *Xenopus* mitotic extract (SUMO-ranGAP1; ref. 19), this protein runs at 90K. **b**, Spindles assemble around chromatin beads that have been incubated for 90 min in M-phase extract supplemented with rhodamine-labelled tubulin. Microtubules are labelled with rhodamine; bead-associated DNA is labelled with Hoechst dye. **c, d**, Addition of 7-fold excess RanT24N (**c**) blocked spindle formation, as did addition of 20-fold excess ranBP1 and 10 μ M ranGAP (**d**). **e**, Centrosome-induced microtubule nucleation in control M-phase extract. **f**, This nucleation was not inhibited by addition of 7-fold excess RanT24N. Scale bar, 25 μ m.

To determine whether RCC1 function is required for inducing spindle formation, we added RanT24N to the extract, together with the chromatin beads. This completely inhibited microtubule nucleation, even around very large bead clusters (Fig. 2: compare control, b, with c; no microtubules are seen on 71 clusters of more than 15 beads in the presence of RanT24N). Microtubule nucleation was also inhibited when excess ranBP1 and ranGAP were added to the extract, together with the chromatin beads, to reconvert any Ran-GTP generated to Ran-GDP (Fig. 2d: no spindles were seen on 65 clusters of more than 15 beads, although 19 of 37 clusters of more than 30 beads had short internal microtubules). This indicates that both RCC1 activity and Ran-GTP are required for microtubule assembly around chromatin beads. RanT24N had no inhibitory effect on the nucleation of microtubules by centrosomes in M-phase extracts (Fig. 2e, f: there was no reduction in aster number or detectable change in aster morphology), demonstrating that Ran-GTP generation by RCC1 is not essential for microtubule nucleation by purified centrosomes.

To confirm that local RCC1 activity on chromatin beads is required for spindle assembly, we added excess RCC1 to the extract in order to decrease the differential between chromatin-bound and unbound RCC1. This treatment caused a partial uncoupling of microtubule assembly from chromatin beads (Fig. 3a, b). Results were similar when RanQ69L-GTP (Fig. 3c) or wild-type Ran loaded with GTP- γ S was added instead (data not shown). Adding wild-type Ran-GTP to the extract in 10-fold excess over the endogenous

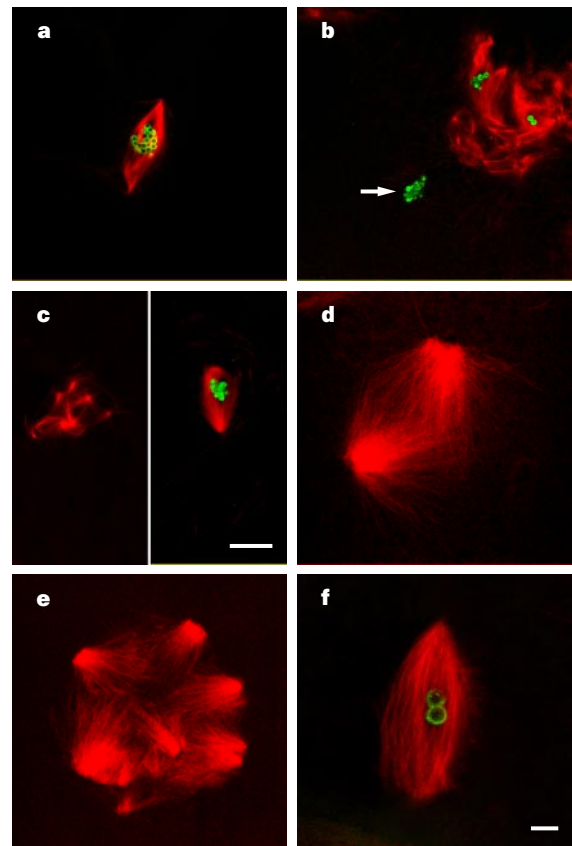


Figure 3 A uniform increase in RanGTP concentration uncouples microtubule assembly from chromatin beads. **a**, A control spindle assembled in the presence of buffer alone. **b, c**, In the presence of 20-fold excess RCC1 (**b**), or 5-fold excess RanQ69LGTP (**c**). **d, e**, Addition of 10-fold excess wild-type Ran loaded with GTP induced the formation of asters that fuse into bipolar (**d**), or multipolar (**e**) structures without chromatin beads. **f**, Addition of 5-fold excess wild-type Ran-GTP did not promote free microtubule assembly but induced the formation of spindles around as few as two chromatin beads. Scale bars: **a-c**, 25 μ m; **d-f**, 5 μ m.

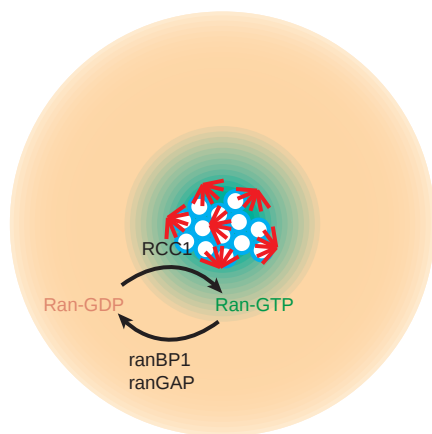


Figure 4 A model for the induction of local microtubule growth. Chromatin-associated RCC1 (blue) locally increases the rate of exchange of GTP for GDP on soluble Ran (Ran-GTP, green; Ran-GDP, orange). Above a threshold concentration of Ran-GTP, microtubule polymerization is induced (red).

concentration¹⁴ also resulted in partial uncoupling of microtubule assembly from chromatin beads (Fig. 3d, e): 41% of bead clusters were now not associated with microtubules, whereas 94.7% of all microtubule assemblies were not associated with beads. Increasing the overall Ran-GTP concentration in the extract therefore gave rise to competition between microtubule nucleation at random sites and normal microtubule assembly around chromatin beads. When wild-type Ran-GTP was added in 5-fold excess, no free nucleation was observed, but as few as two beads could induce spindle formation (Fig. 3f), a situation that is never seen in control extracts^{6,21}. All 94 clusters consisting of more than three beads gave rise to spindle-like structures. We suggest that a small general increase in free Ran-GTP makes it possible for less chromatin-associated RCC1 activity to generate enough Ran-GTP to induce local microtubule nucleation.

In summary, RCC1 was enriched on the mitotic chromatin beads. Inhibiting RCC1 from producing Ran-GTP by RanT24N, or increasing the rate of reconversion of Ran-GTP generated by RCC1 activity back to the GDP-bound state by using excess ranBP1 and ranGAP, prevents chromatin from nucleating microtubules. Spindle assembly could also be affected by treatments such as addition of excess RCC1 or RanQ69L-GTP, which produce high but delocalized Ran-GTP concentrations. These treatments led to induction of microtubule formation and assembly into organized structures, but uncoupled these events from the requirement for added chromatin.

A model based on our data is summarized in Fig. 4. The high local concentration of chromatin-associated RCC1 gives rise to Ran-GTP. The presence of soluble ranBP1 and ranGAP in the extract ensures that the level of Ran-GTP is kept low outside the immediate vicinity of the chromatin. The antagonistic effects of the chromatin-associated exchange factor and the soluble GAP result in high Ran-GTP concentration only around chromatin. Ran-GTP in turn induces local microtubule assembly which, together with the action of microtubule-based motors and probably other proteins^{1,6,21,22}, results in the organization of a bipolar spindle. Our results provide a molecular basis for the effect of chromosomes on local microtubule growth that is seen in dividing plant cells and meiotic animal cells, and we propose that local Ran-GTP production around chromosomes is required for spindle assembly in most living systems. □

Methods

Expression of recombinant proteins. The expression and purification of histidine-tagged human Ran, RanQ69L, RanT24N, ranBP1 and

Schizosaccharomyces pombe Rna1 have been described^{16,23}. The coding region of RCC1 was generated by the polymerase chain reaction (PCR) using the pETRCC1 construct (gift from C. Klebe) as template and cloned between the *Sall* and *Bam*H1 sites of pQE30 (Qiagen). Ran proteins were loaded with GTP where indicated as described²⁴.

Preparation of CSF (M-phase) extract and chromatin beads. CSF-arrested 10,000g *Xenopus laevis* egg extracts (meiotic extracts) were prepared as described²⁵. DNA and chromatin beads were prepared as described^{6,21} except that DNA beads contained more DNA (0.7 pg DNA per bead) and chromatin beads were assembled in the presence of 1 μ M nocodazole in the chromatin assembly extract to avoid any premature polymerization of tubulin.

Freezing of CSF extracts and chromatin beads. Liquid ethane was prepared as for cryo-EM grid freezing²⁶. Extracts and chromatin beads were frozen immediately using thin-wall PCR tubes that were plunged carefully into liquid ethane. Frozen aliquots were stored in liquid nitrogen and thawed rapidly before use.

Experiments and visualization of microtubule figures. Reactions consisted of 20 μ l CSF extract, 0.2 μ l rhodamine-labelled tubulin⁶ and proteins as specified in the text. Endogenous concentrations of Ran, RCC1 and ranBP1 in these extracts were roughly 2, 0.2 and 0.4 μ M, respectively^{14,27,28}. Samples were incubated at 20 °C for 30 min; for 30 min in the presence of 1 μ l of purified centrosomes²⁹; or for 45–90 min in the presence of 0.3 μ g chromatin beads (thawed and collected immediately before use). In the controls, equivalent amounts of buffer were added without the corresponding proteins. Samples were simply fixed and squashed⁶ or spun down onto coverslips³⁰. Image acquisition and analysis was on a Leica confocal microscope.

Western-blot analysis. Chromatin beads were collected, washed twice in 25 mM sodium phosphate, pH 7.5, 100 mM NaCl, and boiled in SDS buffer. Proteins from *Xenopus* extract were precipitated using 66% ethanol, washed, resuspended in SDS sample buffer and boiled. About 10% of extract (2 μ l) and 100% of the bead fraction were used. Human RCC1 and ranBP1 and yeast Rna1 served as positive controls. Proteins were applied on a 12% SDS-polyacrylamide gel, transferred to nitrocellulose (Schleicher and Schuell) and probed with antibodies specific for RCC1 (1:1,000), ranBP1 (1:5,000) and ranGAP (1:1,000)^{18–20}.

Received 5 May; accepted 20 May 1999.

- Walczak, C. E., Vernos, I., Mitchison, T. J., Karsenti, E. & Heald, R. A model for the proposed roles of different microtubule-based motor proteins in establishing spindle bipolarity. *Curr. Biol.* **8**, 903–913 (1998).
- Hoyt, M. A. & Geiser, J. R. Genetic analysis of the mitotic spindle. *Annu. Rev. Genet.* **30**, 7–33 (1996).
- Karsenti, E., Newport, J. & Kirschner, M. The respective roles of centrosomes and chromatin in the conversion of microtubule arrays from interphase to metaphase. *J. Cell Biol.* **99**, 47s–54s (1984).
- Karsenti, E., Newport, J., Hubble, R. & Kirschner, M. Interconversion of metaphase and interphase microtubule arrays, as studied by the injection of centrosomes and nuclei into *Xenopus* eggs. *J. Cell Biol.* **98**, 1730–1745 (1984).
- Hyman, A. & Karsenti, E. The role of nucleation in patterning microtubule networks. *J. Cell Sci.* **111**, 2077–2083 (1998).
- Heald, R. et al. Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382**, 420–425 (1996).
- Koepp, D. M. & Silver, P. A. A GTPase controlling nuclear trafficking: running the right way or walking randomly. *Cell* **87**, 1–4 (1996).
- Görlisch, D. Transport into and out of the cell nucleus. *EMBO J.* **17**, 2721–2727 (1998).
- Mattaj, J. W. & Englmeier, L. Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* **67**, 265–306 (1998).
- Nakamura, M. et al. When overexpressed, a novel centrosomal protein, RanBPM, causes ectopic microtubule nucleation similar to γ -tubulin. *J. Cell Biol.* **143**, 1041–1052 (1998).
- Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A. & Ponstingl, H. RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc. Natl Acad. Sci. USA* **91**, 2587–2591 (1994).
- Klebe, C., Bischoff, F. R., Ponstingl, H. & Wittinghofer, A. Interaction of the nuclear GTP-binding protein Ran with its regulatory proteins RCC1 and RanGAP1. *Biochemistry* **34**, 639–647 (1995).
- Ullrich, O., Reinsch, S., Urbé, S., Zerial, M. & Parton, R. G. Rab11 regulates recycling through the pericentriolar recycling endosome. *J. Cell Biol.* **135**, 913–924 (1996).
- Palacios, I., Weis, K., Klebe, C., Mattaj, J. W. & Dingwall, C. Ran/TC4 mutants identify a common requirement for snRNP and protein import into the nucleus. *J. Cell Biol.* **133**, 485–494 (1996).
- Görlisch, D., Panté, N., Kutay, U., Aebi, U. & Bischoff, F. R. Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.* **15**, 5584–5594 (1996).
- Izaurrealde, E., Kutay, U., von Kobbe, C., Mattaj, J. W. & Görlisch, D. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *EMBO J.* **16**, 6535–6547 (1997).
- Dasso, M., Seki, T., Azuma, Y., Ohba, T. & Nishimoto, T. A mutant form of the Ran/TC4 protein disrupts nuclear function in *Xenopus laevis* egg extracts by inhibiting the RCC1 protein, a regulator of chromosome condensation. *EMBO J.* **13**, 5732–5744 (1994).
- Dasso, M., Nishitani, H., Kornbluth, S., Nishimoto, T. & Newport, J. W. RCC1, a regulator of mitosis, is essential for DNA replication. *Mol. Cell. Biol.* **12**, 3337–3345 (1992).
- Saitoh, H., Pu, R., Cavenagh, M. & Dasso, M. RanBP2 associated with Ubc9p and a modified form of RanGAP1. *Proc. Natl Acad. Sci. USA* **94**, 3736–3741 (1997).
- Pu, R. T. & Dasso, M. The balance of RanBP1 and RCC1 is critical for nuclear assembly and nuclear transport. *Mol. Biol. Cell* **8**, 1955–1970 (1997).

21. Heald, R., Tournebise, R., Habermann, A., Karsenti, E. & Hyman, A. Spindle assembly in *Xenopus* egg extracts: respective roles of centrosomes and microtubule self-organization. *J. Cell Biol.* **138**, 615–628 (1997).
22. Andersen, S. S. *et al.* Mitotic chromatin regulates phosphorylation of Stathmin/Op18. *Nature* **389**, 640–643 (1997).
23. Kutay, U., Izaurralde, E., Bischoff, F. R., Mattaj, J. W. & Görlich, D. Dominant-negative mutants of importin- β block multiple pathways of import and export through the nuclear pore complex. *EMBO J.* **16**, 1153–1163 (1997).
24. Weis, K., Dingwall, C. & Lamond, A. I. Characterization of the nuclear protein import mechanism using Ran mutants with altered nucleotide binding specificities. *EMBO J.* **15**, 7120–7128 (1996).
25. Murray, A. in *Xenopus laevis: Practical Uses in Cell and Molecular Biology* (eds Kay, B. K. & Peng, H. B.) 581–605 (Academic, San Diego, 1991).
26. Lepault, J. & Dubochet, J. Electron microscopy of frozen hydrated specimens: preparation and characteristics. *Meth. Enzymol.* **127**, 719–730 (1986).
27. Nicolás, F. *et al.* *Xenopus* Ran-binding protein 1: molecular interactions and effects on nuclear assembly in *Xenopus* egg extracts. *J. Cell Sci.* **110**, 3019–3030 (1997).
28. Clarke, P. R., Klebe, C., Wittinghofer, A. & Karsenti, E. Regulation of Cdc2/cyclin B activation by Ran, a Ras-related GTPase. *J. Cell Sci.* **108**, 1217–1225 (1994).
29. Domínguez, J. E. *et al.* A protein related to brain microtubule-associated protein MAP1B is a component of the mammalian centrosome. *J. Cell Sci.* **107**, 601–611 (1994).
30. Wittmann, T., Boleti, H., Antony, C., Karsenti, E. & Vernos, I. Localization of the kinesin-like protein xklp2 to spindle poles requires a leucine zipper, a microtubule-associated protein, and dynein. *J. Cell Biol.* **143**, 673–685 (1998).

Acknowledgements. We thank M. Dasso for antibodies and for discussing unpublished data; C. Klebe and A. Wittinghofer for pETRCC1; B. Soennichsen, M. Zerial, M. Fornerod, M. Ohno, G.-J. Arts and D. Bilbao-Cortes for Rab and Ran proteins; T. Surrey for participating in the 'ethane trick'; and T. Wittmann and our colleagues for discussion and encouragement. G.G. was supported by a fellowship from the MURST Italian Ministry of Public Education and by an EMBO short-term fellowship, and O.J.G. was supported by an HFSP award to I.W.M.

Correspondence and requests for materials should be addressed to I.W.M. (e-mail: mattaj@embielheidelberg.de).

Orientation of DNA replication establishes mating-type switching pattern in *S. pombe*

Jacob Z. Dalggaard & Amar J. S. Klar

Gene Regulation and Chromosome Biology Laboratory, ABL-Basic Research Program, NCI Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA

The fission yeast *Schizosaccharomyces pombe* normally has haploid cells of two mating types, which differ at the chromosomal locus *mat1*. After two consecutive asymmetric cell divisions, only one in four 'grand-daughter' cells undergoes a 'mating-type switch', in which genetic information is transferred to *mat1* from the *mat2-P* or *mat3-M* donor loci^{1–4}. This switching pattern probably results from an imprinting event at *mat1* that marks one sister chromatid in a strand-specific manner^{3–5}, and is related to a site-specific, double-stranded DNA break at *mat1*^{6,7}. Here we show that the genetic imprint is a strand-specific, alkali-labile DNA modification at *mat1*. The DNA break is an artefact, created from the imprint during DNA purification. We also propose and test the model that *mat1* is preferentially replicated by a centromere-distal origin(s), so that the strand-specific imprint occurs only during lagging-strand synthesis. Altering the origin of replication, by inverting *mat1* or introducing an origin of replication, affects the imprinting and switching efficiencies in predicted ways. Two-dimensional gel analysis confirmed that *mat1* is preferentially replicated by a centromere-distal origin(s). Thus, the DNA replication machinery may confer different developmental potential to sister cells.

As strains that are deleted from donor loci are alive despite having a double-stranded DNA break (DSB) at *mat1* (ref. 8), we investigated whether the break is an artefact of DNA purification. We compared the DSB level of DNA gently extracted from cells embedded in agarose with that of DNA purified by the standard method. As shown originally⁶, approximately one-quarter of the chromosomal DNA is broken at *mat1* when the DNA is purified from *h*⁹⁰ (wild-type SP976) cells by the standard method (Fig. 1a). However, there was no DSB in DNA purified in agarose. We

analysed this discrepancy further by denaturing the DNA. To simplify the interpretation of autoradiograms, we used strains without donor cassettes, thereby eliminating hybridization signals for *mat2-P* and *mat3-M* (ref. 8). Again, no break could be detected in *HindIII*-digested DNA displayed on the native gel (SP714; Fig. 1b). Denaturing with formamide and formaldehyde treatment indicated that both DNA strands were intact. However, denaturing with sodium hydroxide produced a break at *mat1* in a fraction of the DNA (SP714), but not in the *mat1-Msmt-0* (JZ108) strain that lacks the imprint. Furthermore, hybridization with strand-specific probes showed that only the upper strand of the DNA was labile (Fig. 1c). The labile strand is the strand in which the break was mapped⁷. These results biochemically establish the existence of a strand-specific, alkali-labile modification(s) in the DNA that correlates with the switching ability. The level of imprinting/DSB formation was affected by growth conditions; for example, a ~30% or ~10% break level was detected (data not shown) when the cells were grown in poor (PMA⁺) or rich (YEA) medium⁹, respectively.

The strand-specificity of the imprint could be achieved through

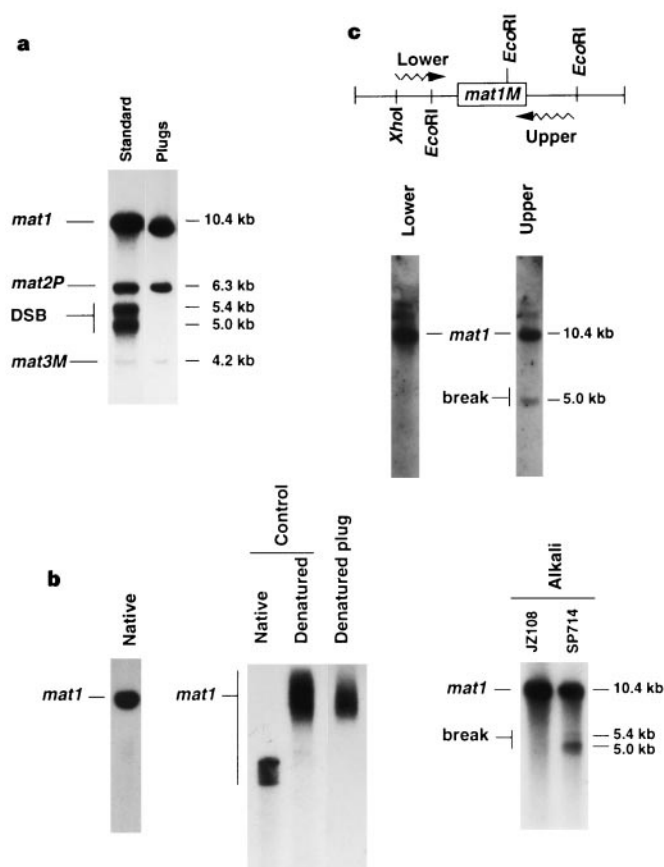


Figure 1 Characterization of the imprint. Southern analysis of *HindIII*-digested DNA using a 10.4-kb *mat1-P* *HindIII* fragment as the probe. DNA was purified by the standard method⁹ when indicated and otherwise by the 'plug' method of embedding the cells in agarose²³. Cultures were grown in rich (YEA) medium for panel **c**, and otherwise in sporulation (PMA⁺) medium. **a**, The DSB is an artefact of DNA purification. The 5.4 and 5.0-kb fragments result from the DSB in the 10.4-kb *mat1* fragment. The *mat2-P* (6.3 kb) and *mat3-M* (4.2 kb) bands result from hybridization to the *mat1* probe as they share cassette homology (Fig. 2a). **b**, Alkali treatment, but not formaldehyde denaturation, creates a break at *mat1*. As a control for the denaturation of the sample run on the formaldehyde gel, analysis of native and denatured *mat1P* *HindIII* fragment, derived from a plasmid amplified in *Escherichia coli*, as shown. **c**, The break is strand-specific. Alkali-denatured DNA was probed with the indicated strand-specific RNA probes (zigzag lines). The probe marked 'upper' hybridizes to the upper strand and the other gives a signal with the 'lower' strand.

the lagging- and leading-stranded nature of DNA replication. We therefore propose that the *mat1* cassette is replicated by a distally-located origin(s) and that only the lagging strand (red, Fig. 2a, b) is imprinted (star, Fig. 2b) during its synthesis. Furthermore, the cell that inherits the imprinted chromosome becomes switchable. Leading-strand replication (blue, Fig. 2b) in this cell is postulated to result in a transient break due to the inherited modification in the template (hypothetical intermediate indicated by shading, Fig. 2b). This transient break initiates a recombinational interaction with one of the intact donor loci, resulting in a switch. This model is supported by the finding that *swi7⁺*, a gene essential for producing the DSB and for switching^{5,10}, encodes the catalytic subunit of DNA polymerase- α ¹¹. DNA polymerase- α is predominantly involved in lagging-strand synthesis. A testable feature of the model is that the efficiency of imprinting and switching should be sensitive to the replication orientation of *mat1* DNA.

To test this model, an origin (A) was introduced on either side of *mat1*. We used a 2.8-kilobase (kb) *PvuII* fragment containing the *S. pombe* bidirectionally replicating origin (pYAm6.2)¹². The origin A, substituted for the 0.8-kb *EcoRI* *mat1*-(centromere)proximal fragment (JZ167; Fig. 3), reduced both switching (61%) and DSB formation (71%). Two-dimensional gel analysis detected low origin-A activity (indicated by the bubble-arc) corresponding to the observed effect on switching (Fig. 4). We also tested the effect of another proximally placed origin, *ars1* (ref. 13). This strain showed similarly reduced sporulation (72%) and DSB formation (80%) (data not shown). However, as expected, origin A introduced into an *h⁹⁰* strain (SP976) at the *mat1*-distal *SspI* site by using the *Saccharomyces cerevisiae* *LEU2* gene as a selectable marker (JZ142) did not significantly affect switching (Fig. 3). These data are

consistent with our model.

A more critical prediction of the model is that an inversion of the 1.4-kb *SspI* fragment, which primarily contains the *mat1* gene, should greatly reduce switching. This fragment contains sufficient sequences for imprinting and switching^{3,4}. As expected, the inversion resulted in an almost non-switching phenotype, indicated by a very low rate (0.3%) of sporulation, lack of iodine-vapour staining of colonies and absence of the DSB (JZ148; Fig. 3a, b).

If the greatly reduced switching of the inverted *mat1* cassette is due to interchange with respect to lagging- and leading-strand replication of DNA chains, introduction of an origin proximal to *mat1* should partially restore switching. For full restoration, the introduced origin has to fire in all cells and the hypothesized replication of *mat1* coming from the other direction (Fig. 2b and see below) has to be blocked. Introduction of origin A proximal to *mat1* increased the sporulation rate to 11%, but the amount of the DSB was below the level of reliable quantification (JZ169; Fig. 3a, b). The increased low-level switching corresponds to the low activity of this origin (Fig. 4). Using genetic experiments, we found that this increased level of switching depended on the *swi1⁺*, *swi3⁺* and *swi7⁺* genes (data not shown), indicating that increased switching may occur through the standard pathway of mating-type interconversion. Similarly, placing another origin, *ars1* (ref. 13), at the same position partially restored sporulation and imprinting at *mat1* (JZ179; Fig. 3).

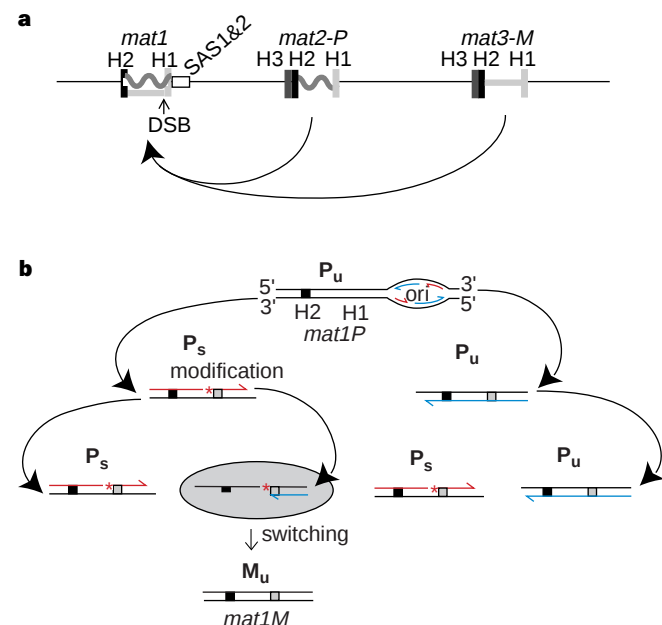


Figure 2 Mating-type region and the pattern of switching of mating types. **a**, The mating-type region of chromosome II^{19,20}. *mat2-P* and *mat3-M* encode the P and M information, respectively. *SAS1* and *SAS2* indicate *cis*-acting sequences essential for imprinting and switching²¹; H1, H2 and H3 are homology boxes that flank allele-specific sequences; DSB indicates the site of the double-strand break. Several unlinked genes (*swi1⁺*, *swi3⁺* or *swi7⁺*) and a *cis*-acting sequence defined by the *mat1-Msm10* allele are involved in the formation and/or maintenance of the imprinting event^{5,10,12,22}. Arrows indicate replacement of *mat1* with a copy of one of the unexpressed *mat2* or *mat3* loci, resulting in a mating-type switch. **b**, The pattern of switching in cell pedigrees superimposed on the 'replication-orientation model'. The subscripts 'u' and 's' designate unswitchable and switchable cells, respectively. Arrows indicate the inheritance of a particular DNA chain by a daughter cell. 'ori' indicates the origin of replication. See text for details.

a		Switching rate or (spo %)	DSB	Staining
SP976		100 %	100 %	
JZ142		98 %	91 %	
JZ167		61 %	71 %	
JZ148		(0.3 %)	undetectable	
JZ169		(11 %)	very low	
JZ179		(4 %)	7 %	

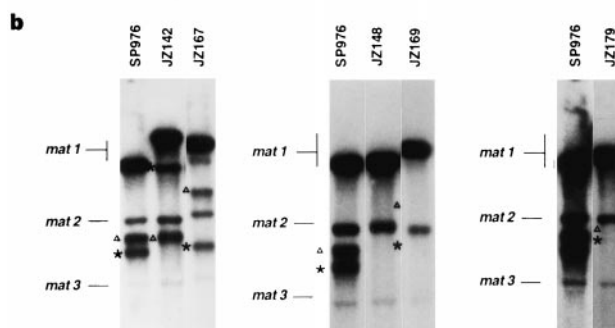


Figure 3 Analysis of rearrangements. **a**, Effect of rearrangements on switching. Names and schematic drawings of the changes made in the *mat1* region are indicated. oriA, pYAm6.2; oriB, *ars1*. The switching rates and sporulation frequencies (in parentheses) are presented as the percentage of values ordered for the parental SP976 control strain. Iodine vapours stain efficiently switching (spore-containing) colonies uniformly black, whereas non-switching/non-sporulating colonies stain yellow²⁴. **b**, Effect of rearrangements on the DSB level. The proximal and distal fragments generated by the DSB are indicated by a triangle and a star, respectively.

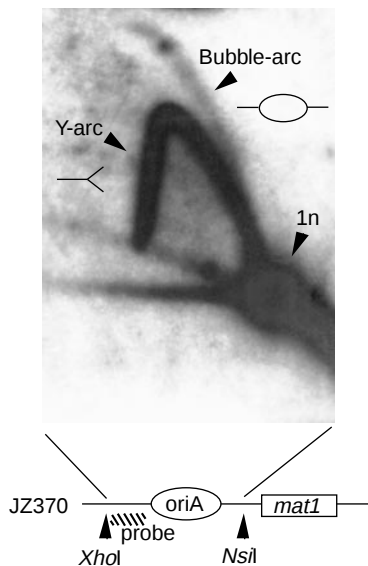


Figure 4 Assessment of origin activity. Standard two-dimensional gel analysis¹⁴ was performed on strain JZ370. JZ370 is derived from JZ167 by crossing in the *swi3*⁻ mutation, which reduces switching¹⁰. Signals of switching intermediates otherwise interfere with the quantification of origin activity. The analysis detected a very strong Y-arc and a weak (~10% of total) bubble-arc. A Y-arc is observed when a fragment is replicated by an origin located outside the fragment, whereas replication initiated from the origin produces a bubble-arc. 1n, linear unreplicated fragment.

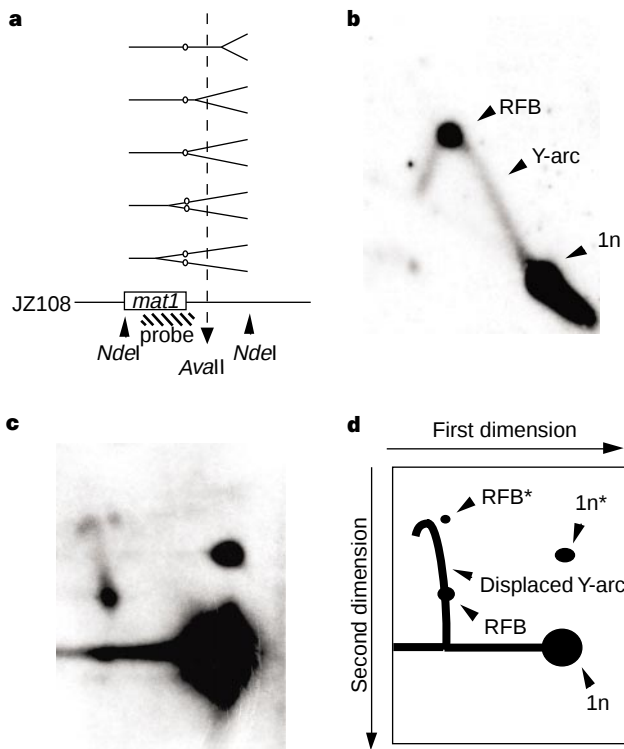


Figure 5 Determination of replication orientation at *mat1* of strain JZ108. **a**, The region contained in the *NdeI* fragment was analysed. A schematic drawing of the replication fork initiated from *mat1*-distal origin is shown. The position of an observed replication fork barrier (RFB) is indicated by a circle. **b**, Standard two-dimensional gel analysis. **c**, Fork direction experiment. The indicated *Avall* digestion was done before separating the DNA in the second dimension. **d**, Diagram of interpretation of the autoradiogram in **c**. 1n, linear unreplicated fragment; 1n*, undigested *NdeI* fragment; RFB, replication fork barrier; RFB*, intermediate where only one of the two *Avall* sites in the replicated arms has been digested.

To determine directly the orientation of *mat1* replication predicted in the model (Fig. 1b), we used the method described in ref. 14. We used the non-switching *mat1*-*Msm*0 strain (JZ108), as the presence of switching intermediates (unpublished observations) in wild-type cells complicates the interpretation of the results. The *mat1* fragment lacks an origin, as only a so-called 'Y-arc' is observed by two-dimensional gel analysis (Fig. 5b). Additionally, we found a replication fork barrier (pause site) in the region. Using the method for determining the direction of fork movement¹⁴, we saw a displaced arc when we used a probe specific to the proximal part of the *NdeI* fragment (Fig. 5c, d). This showed that most *mat1* DNA is replicated from an origin(s) located distal to it. This experiment also mapped the replication fork barrier to the region just distal to the *mat1* cassette at the H1 domain (Fig. 5a, c). The significance of this pause site is not known.

Here we have proposed and tested a replication-orientation model that explains the underlying molecular mechanism for the asymmetric switching pattern of *S. pombe*. The model indicates that the pattern may be due to a heritable strand-specific imprinting event that modifies the strand synthesized only during lagging-strand replication, and that *mat1* switching occurs during leading-strand replication of an imprinted chromosome. This proposed mechanism of induction of switching also explains the 3:1 gene conversion observed at *mat1* during meiosis⁸. The model is based on our discovery that the previously described imprinting event^{3,4} is a strand-specific, alkali-labile modification of the DNA. The alkali-labile character of the modification and the involvement of DNA polymerase- α in formation of the imprint may mean that the imprint is an RNA primer left from DNA replication that has been ligated to form a continuous DNA-RNA strand. Such a modification could also explain why the standard DNA purification procedure artefactually converts the imprint into a DSB, owing to RNase A treatment combined with an unintentional shearing. Other work has shown that the imprint at *mat1* is instead a strand-specific nick¹⁵. This discrepancy could be explained if the base-labile modification we have discovered is converted into a nick during the heat-treatment of the DNA in that study¹⁵. To our knowledge, this is the first example in which the orientation of DNA replication establishes the asymmetric developmental potential of two sister cells. In this regard, the mating-type switching pedigree resembles the stem-cell lineage pattern of higher eukaryotes^{2,16}. □

Methods

***S. pombe* strains.** Strains were constructed using standard cloning and transformation methods⁹ from wild-type *h*⁹⁰ strain SP976 (*h90*, *ura4-D18*, *leu1-32*, *ade6-M210*), SP714 (*mat1M*, *mat2,3Δ::LEU2*, *leu1-32*, *ura4-D18*, *ade6-M216*) and JZ108 (*mat1*-*Msm*0, *mat2,3Δ::LEU2*, *leu1-32*, *ade6-M210*, *his2*).

Denaturation of DNA embedded in agarose. DNA in plugs was denatured by soaking it in 65% formamide, 8% formaldehyde solution for 15 min, followed by incubation at 65 °C for 5 min¹⁷. For the alkali procedure, the DNA was denatured by incubation in 40 mM NaOH, 2 mM EDTA for 2 h¹⁸.

Rates of switching. For efficiently switching strains, the rate of switching was determined from ~100 single-cell pedigrees as described¹. For strains that switched inefficiently, the percentage of sporulation was determined instead (~300 cells+asci counted).

Quantification of the DSB. DNA was purified from cells grown in sporulation media using the DNAzol kit (GibcoBRL) plus an additional step of phenol extraction and ethanol precipitation. This method gave reproducible levels of the DSB. The level of the DSB, determined by phosphorimager analysis (Molecular Dynamics), is indicated relative to that of the wild-type strain (SP976). For comparison between different strains, the *mat2* and *mat3* signals were used as internal loading controls.

Two-dimensional gel electrophoresis. Electrophoresis was performed in a 0.4% agarose gel in the first dimension, the second separation was performed in a 1.1% agarose. To determine the fork direction, the second separation was performed in 1.4% agarose.

Received 24 December 1998; accepted 17 May 1999.

- Miyata, H. & Miyata, M. Mode of conjugation in homothallic cells of *Schizosaccharomyces pombe*. *J. Gen. Appl. Microbiol.* **27**, 365–371 (1981).
- Egel, R. & Eie, B. Cell lineage asymmetry in *Schizosaccharomyces pombe*: unilateral transmission of a high-frequency state of mating-type switching in diploid pedigrees. *Curr. Genet.* **12**, 429–433 (1987).
- Klar, A. J. S. Differentiated parental DNA strands confer developmental asymmetry on daughter cells in fission yeast. *Nature* **326**, 466–470 (1987).
- Klar, A. J. S. The developmental fate of fission yeast cells is determined by the pattern of inheritance of parental and grandparental DNA strands. *EMBO J.* **9**, 1407–1415 (1990).
- Klar, A. J. S. & Bonaduce, M. J. The mechanism of fission yeast mating-type interconversion: evidence for two types of epigenetically inherited chromosomal imprinted events. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 457–465 (1993).
- Beach, D. H. Cell type switching by DNA transposition in fission yeast. *Nature* **305**, 682–687 (1983).
- Nielsen, O. & Egel, R. Mapping the double-strand breaks at the mating-type locus in fission yeast by genomic sequencing. *EMBO J.* **8**, 269–276 (1989).
- Klar, A. J. S. & Miglio, L. M. Initiation of meiotic recombination by double-strand DNA breaks in *S. pombe*. *Cell* **46**, 725–731 (1986).
- Moreno, S. A., Klar, A. J. S. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
- Egel, R., Beach, D. H. & Klar, A. J. S. Genes required for initiation and resolution steps of mating-type switching in fission yeast. *Proc. Natl Acad. Sci. USA* **81**, 3481–3485 (1984).
- Singh, J. & Klar, A. J. S. DNA polymerase- α is essential for mating-type switching in fission yeast. *Nature* **361**, 271–273 (1993).
- Caddle, M. S. & Calos, M. P. Specific initiation at an origin of replication from *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **14**, 1796–1805 (1994).
- Clyne, R. K. & Kelly, T. J. Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* **14**, 6348–6357 (1995).
- Brewer, B. J., Lockshon, D. & Fangman, W. L. The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell* **71**, 267–276 (1992).
- Arcangeli, B. A site- and strand-specific DNA break confers asymmetric switching potential in fission yeast. *EMBO J.* **17**, 4503–4510 (1998).
- Klar, A. J. S. Lineage-dependent mating-type transposition in fission and budding yeast. *Curr. Opin. Genet. Dev.* **3**, 745–751 (1993).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd ed (Cold Spring Harbor Laboratory Press, New York, 1989).
- MacDonell, M. W., Simon, M. W. & Studier, F. W. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* **110**, 119–146 (1977).
- Egel, R. in *Molecular Biology of the Fission Yeast* (eds Nasim, A., Young, P. & Johnson, B. F.) 31–74 (Academic, San Diego, 1989).
- Klar, A. J. S., Ivanova, A. V., Dalgaard, J. Z., Bonaduce, M. J. & Grewal, S. I. S. in *Epigenetics* (ed. Cardew, G.) 87–103 (Novartis Foundation Symposium 214, Wiley, Chichester, 1998).
- Arcangeli, B. & Klar, A. J. S. A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient *mat1* switching in *Schizosaccharomyces pombe*. *EMBO J.* **10**, 3025–3032 (1991).
- Styrkarsdottir, U., Egel, R. & Nielsen, O. The *smt-0* mutation which abolishes mating-type switching in fission yeast is a deletion. *Curr. Genet.* **23**, 184–186 (1993).
- Gerring, S. L., Connelly, C. & Hieter, P. in *Guide to Yeast Genetics and Molecular Biology* (eds Fink, G. & Guthrie, C.) 57–76 (Academic, San Diego, 1991).
- Bresch, C., Muller, G. & Egel, R. Genes involved in meiosis and sporulation of a yeast. *Mol. Gen. Genet.* **102**, 301–306 (1968).

Acknowledgements. We thank our colleagues at ABL for suggestions and helpful discussions; B. Brewer, M. Hoant and S. Hunt for help with the analysis of replication intermediates; A. Arthur for editorial suggestions; and R. Frederickson for the art work. This work was sponsored by the Danish Natural Science Research Council (J.Z.D.) and the National Cancer Institute, Department of Health and Human Services, under contract with ABL. The contents of this article do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Correspondence and requests for materials should be addressed to A.K. (e-mail: klar@mail.ncicrf.gov).

Single kinesin molecules studied with a molecular force clamp

Koen Visscher^{*†‡}, Mark J. Schnitzer^{*§} & Steven M. Block^{*†§}

^{*} Department of Molecular Biology, [†] Princeton Materials Institute, and [§] Department of Physics, Princeton University, Princeton, New Jersey 08544, USA
[‡] These authors contributed equally to this work.

Kinesin is a two-headed, ATP-driven motor protein that moves processively along microtubules in discrete steps of 8 nm, probably by advancing each of its heads alternately in sequence^{1–4}. Molecular details of how the chemical energy stored in ATP is coupled to mechanical displacement remain obscure. To shed light on this question, a force clamp was constructed, based on a feedback-driven optical trap capable of maintaining constant loads on single kinesin motors⁵. The instrument provides unprecedented resolution of molecular motion and permits

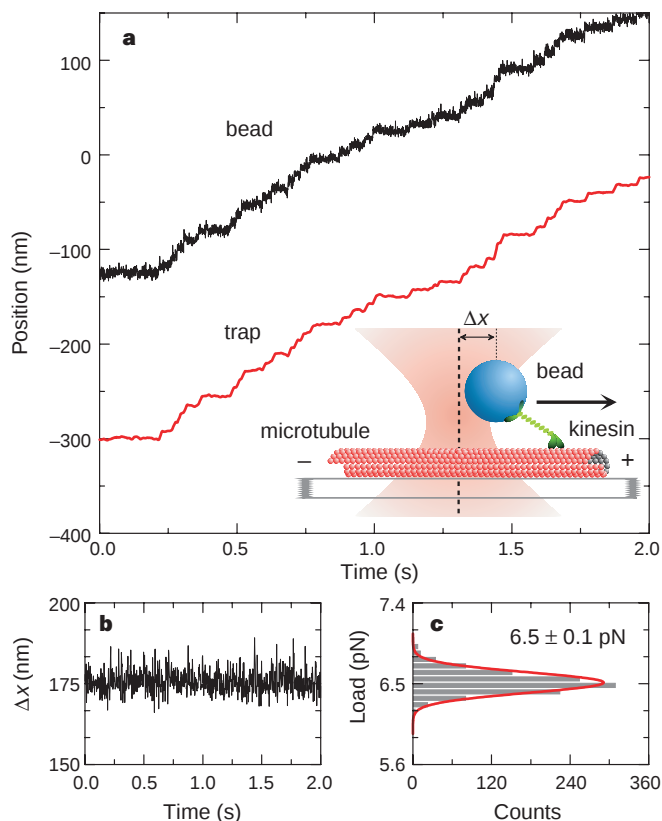


Figure 1 Operation of the force clamp. **a**, Sample record from the force clamp, showing kinesin-driven bead movement and corresponding optical trap displacement (2 mM ATP). Discrete steps of 8 nm are readily apparent. Inset, schematic representation of the motility assay used, showing the experimental geometry (not to scale). The separation between bead and trap was nominally fixed at $\Delta x = 175$ nm. Bead position was sampled at 20 kHz, filtered with a 12-ms boxcar window for the feedback on the trap deflection, and saved unfiltered at 2.0 kHz. **b**, The measured bead-trap separation, Δx , for the record in **a**. **c**, Histogram of the displacements in **b**, converted to force by multiplying by the trap stiffness (0.037 pN nm⁻¹). Solid red line is Gaussian fit to these data, yielding a load of 6.5 ± 0.1 pN (mean $\pm$ s.d.).

mechanochemical studies under controlled external loads. Analysis of records of kinesin motion under variable ATP concentrations and loads revealed several new features. First, kinesin stepping appears to be tightly coupled to ATP hydrolysis over a wide range of forces, with a single hydrolysis per 8-nm mechanical advance. Second, the kinesin stall force depends on the ATP concentration. Third, increased loads reduce the maximum velocity as expected, but also raise the apparent Michaelis–Menten constant. The kinesin cycle therefore contains at least one load-dependent transition affecting the rate at which ATP molecules bind and subsequently commit to hydrolysis. It is likely that at least one other load-dependent rate exists, affecting turnover number. Together, these findings will necessitate revisions to our understanding of how kinesin motors function.

The mechanochemical coupling ratio relates the number of ATP molecules hydrolysed to the distance travelled during a single enzymatic cycle. Determination of this ratio over a range of loads sheds light on the mechanism for movement, and has been considered one of the most pressing challenges in the motility field¹. However, for the microtubule-based molecular motor kinesin, the task is by no means straightforward. Direct comparisons between solution ATPase measurements and motility *in vitro* are only feasible in the absence of external loads, as these cannot be applied in bulk solution. Single-molecule fluorescence methods,

which have been used to correlate motor movement with nucleotide binding and release in the actomyosin system^{6,7}, have proved difficult to realize with kinesin because it has a lower affinity for ATP: concentrations of fluorescent ATP analogues sparse enough to detect as single fluorophores yield negligible rates of movement. Coupling ratios may also be inferred from an analysis of fluctuations in displacement records of single molecules, but this approach requires continuous records over comparatively large distances under invariant loads^{8,9}. Such records are hard to obtain using either conventional laser-based optical traps^{10–12} or microneedles¹³, where loads change continually as motors progress.

We solved the latter difficulty by developing an optical force clamp that can record displacements of single kinesin molecules over hundreds of nanometres while maintaining the average load to within a fraction of a piconewton (Fig. 1). In the assay, a kinesin molecule adsorbed to a silica bead (0.5 μm in diameter) moves processively along an immobilized microtubule³ while an optical trap follows its motion (Fig. 1a). Bead movement is monitored with subnanometre resolution by a quadrant photodiode, which detects scattered light from a fibre-coupled HeNe laser source¹⁴. Because force on the bead is proportional to its displacement from the trap centre¹⁵, load can be maintained by a servo system that dynamically adjusts this separation to some preset value, using digital feedback to control acousto-optical deflectors which move the trap position as required (Fig. 1b, c). In contrast to other feedback-enhanced optical traps developed for isometric force measurements^{16–18}, a force clamp is particularly suited to isotonic displacement measurements⁵. A further benefit of a force clamp is that, unlike other optical trapping systems, displacement data do not require corrections for the series elastic compliance between the bead and motor, hitherto a source of experimental uncertainty amounting to 20% or more^{3,5}. Force clamps therefore provide exceptionally clear molecular records, and individual 8-nm steps can be readily visualized at kilohertz bandwidths, even for high ATP levels (Fig. 1a).

We first used this system to study kinesin velocity as a function of ATP concentration at fixed load. For negligible loads, the velocity, v , obeys Michaelis–Menten kinetics, $v = V_{\text{max}}[\text{ATP}]/([\text{ATP}] + K_m)$, in both microtubule gliding⁴ and bead-based^{8,9} assays. The maximum velocity at saturating ATP levels, V_{max} , is expected to fall with increased load, but how will the apparent Michaelis–Menten constant, K_m , change—if at all? Alternative coupling schemes supply different predictions. Simple loosely coupled models¹⁰ postulate a maximum hydrolysis rate, k_{cat} , that is independent of the applied force, F , with a load-dependent coupling efficiency, $\epsilon(F)$, between ATP hydrolysis and mechanical stepping (with $V_{\text{max}} = \epsilon(F)dk_{\text{cat}}/d$; $d = 8 \text{ nm}$). In practice, the coupling ratio has been shown to be independent of ATP concentration and 1:1 at negligible loads^{8,9}; that is, there is one hydrolysis for every 8 nm advance. Because neither k_{cat} nor k_b , the effective rate of ATP binding (defined as $k_b = k_{\text{cat}}/K_m$), depends on load in these models, it follows that $K_m = k_{\text{cat}}/k_b$ must be load-invariant^{10,19}. In contrast, simple tightly coupled models incorporate a single mechanically sensitive rate that affects k_{cat} but leaves ϵ and k_b unchanged. Under increased load, k_{cat} slows, reducing both K_m and V_{max} .

Experimentally, high loads reduced V_{max} as anticipated. Surprisingly, however, they also raised K_m (Fig. 2), a phenomenon that is inconsistent with either simple model. At the lowest load tested ($1.05 \pm 0.01 \text{ pN}$), V_{max} was $813 \pm 28 \text{ nm s}^{-1}$ and K_m was $88 \pm 7 \mu\text{M}$, comparable to values obtained previously for negligible loads^{4,8,9}. At moderate loads ($3.59 \pm 0.03 \text{ pN}$), V_{max} dropped to $715 \pm 19 \text{ nm s}^{-1}$ while K_m rose to $140 \pm 6 \mu\text{M}$. At the highest load tested ($5.63 \pm 0.06 \text{ pN}$), V_{max} dropped further to $404 \pm 32 \text{ nm s}^{-1}$, while K_m rose to $312 \pm 49 \mu\text{M}$. Because $K_m = k_{\text{cat}}/k_b$, any increase in K_m —given the drop in V_{max} —implies that k_b decreases substantially with increased load. Thus, the application of load has two

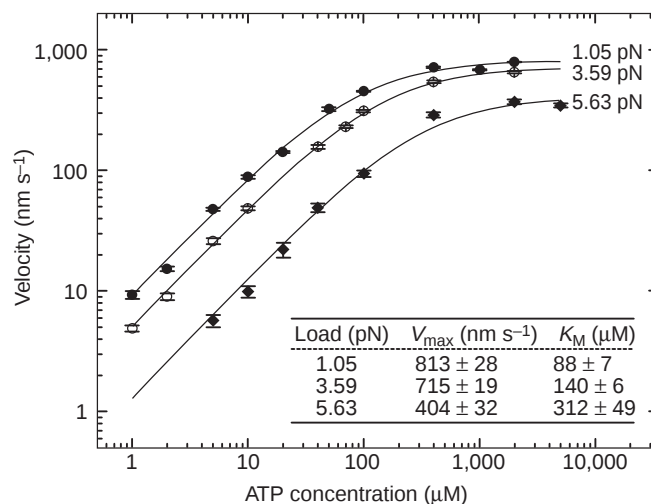
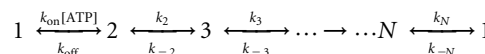


Figure 2 Michaelis–Menten kinetics under load. Double logarithmic plot of the average bead velocity, v (mean $\pm$ s.e.m.), versus ATP concentration for various loads (filled circles, $1.05 \pm 0.01 \text{ pN}$, $N = 11$ –102 runs; open circles, $3.59 \pm 0.03 \text{ pN}$, $N = 8$ –79 runs; diamonds, $5.63 \pm 0.06 \text{ pN}$, $N = 19$ –58 runs). Data were fitted to Michaelis–Menten curves (lines), $v = V_{\text{max}}[\text{ATP}]/([\text{ATP}] + K_m)$. Inset, fit parameters, V_{max} and K_m .

important effects on the kinesin cycle: it lowers the maximal stepping rate, and it also lowers the effective rate of ATP binding.

What causes the effective binding rate to decline under load? For a generalized ATPase pathway involving N intermediate states



there are at least three possible mechanisms. The first is that k_b falls because the second-order ATP binding rate, k_{on} , is reduced by load. The second is that k_b falls because of an increase in the ATP unbinding rate, k_{off} . In either event, the observed drop in V_{max} must arise from some other transition because, for pathways with an irreversible step, such as the kinesin cycle²⁰, neither k_{on} nor k_{off} affects either k_{cat} or ϵ . The third possible mechanism is somewhat more subtle, and involves a reduction in the rate following binding, k_2 . When k_2 falls, the ATP affinity ($k_{\text{on}}/k_{\text{off}}$) is unaffected, but a greater fraction of ATP molecules will unbind before hydrolysis, reducing the effective binding rate (the rate at which ATP molecules are committed to the remainder of the hydrolysis pathway). Because k_2 also affects k_{cat} , lowering k_2 alone could reduce k_{cat} and raise K_m simultaneously, but only if k_{off} were sufficiently fast. (More generally, a change in one or more reversible transitions subsequent to ATP binding but before the first irreversible transition could produce a similar effect, given sufficiently fast reverse rates.) The nearly fourfold increase found for K_m accompanying a drop in speed by a factor of two would require a k_{off} of not less than 650 s^{-1} , which is significantly faster than ATP unbinding rates of 50 – 200 s^{-1} reported in solution studies^{21,22}. However, the latter estimates were determined indirectly and were subject to simplifying assumptions, so their experimental uncertainties remain fairly large. A k_2 -based mechanism therefore ought not be excluded solely on this basis. Notwithstanding, the interpretation we favour is that the contrary behaviour of V_{max} and K_m derives from two independent, mechanically sensitive rates, one of which affects ATP binding through k_{on} and/or k_{off} .

Regardless of the candidate mechanism, the two effects of load on the mechanochemical pathway imply that kinesin force–velocity curves should display different shapes for different ATP concentrations. However, this inference appears to be at odds with several previous studies of force and velocity using either fixed optical

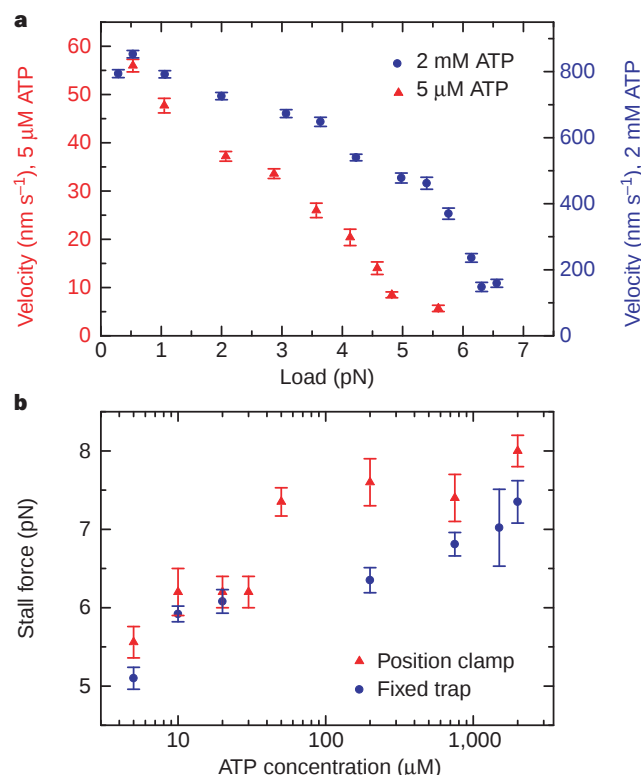


Figure 3 Load dependence of motility. **a**, Average bead velocity, v (mean $\pm$ s.e.m.), versus applied load for fixed ATP concentrations (red triangles, left axis, 5 μ M ATP, $N = 19$ –57; blue circles, right axis, 2 mM ATP, $N = 37$ –87). The velocity point at (5.6 pN, 5 μ M ATP) is likely to represent an overestimate because beads which stalled completely ($v = 0$) were indistinguishable from beads lacking active motors, and so were not included in the analysis. **b**, Stall force (mean $\pm$ s.e.m.) versus ATP concentration, measured either with the position clamp^{17,18} (red triangles) or with a fixed optical trap (blue circles). Stalls had to last a minimum of 2 s to be included in the analysis. Data points represent an average of either 12–29 (position clamp) or 6–70 (fixed trap) stalls.

traps^{10–12} or microneedles¹³, all of which reported nearly linear decreases in velocity with increased load, for both limiting and saturating ATP concentrations. In fact, such linear force–velocity relations formed the basis of an argument for a load-independent K_m (refs 10, 19) which supported a simple, loosely coupled picture¹⁰. However, the earlier studies lacked the improved resolution afforded by the force clamp, and subtle differences in shape were not distinguished within experimental error. Data from all previous studies consistently showed a lower stall force at limiting ATP than at saturating ATP, although the difference in force was insufficient in any given experiment to have been considered significant. Moreover, a greater discrepancy (~ 1 pN) between the stall force at limiting and saturating ATP was observed when the limiting concentration of ATP was ~ 35 -fold lower than the unloaded K_m (ref. 13), as compared to when limiting ATP values were only ~ 6 -fold lower (~ 0.5 pN, refs 10–12). We therefore measured velocity as a function of load in the force clamp at both 2 mM and 5 μ M ATP, choosing the lower ATP level to be >17 -fold below the unloaded K_m to highlight any potential differences.

Force–velocity curves obtained with the force clamp displayed different shapes at limiting and saturating ATP levels (Fig. 3a), and they also showed different stall forces. The 5- μ M curve is nearly linear, with a stall force of ~ 5.5 pN. In contrast, the 2-mM curve is concave down, and supports a greater stall force of ~ 7.0 pN. Strikingly, a 5-pN load leads to a $\sim 40\%$ decline in speed for 2 mM ATP, whereas at 5 μ M ATP an identical load is sufficient to

halt movement almost entirely. These results are consistent with a role for load in lowering the effective rate of ATP binding. At limiting ATP, k_b becomes much lower than k_{cat} , and therefore the stall force becomes independent of the latter. However, at 2 mM ATP, the stall force probably depends on both k_b and k_{cat} , as K_m rises beyond ~ 300 μ M at loads of 7 pN or more. Although the stall force may depend on additional factors at all ATP levels, stall forces are expected to rise continuously from ~ 5.5 pN to ~ 7.0 pN as the ATP concentration is raised.

To further test this supposition, we measured the ATP dependence of kinesin stall force by two additional methods, using either an optical position clamp or a fixed trap. The position clamp¹⁷, based on an optical trapping interferometer, is well suited to isometric measurements of stall force¹⁶, and was used previously to study force generation by individual RNA polymerase molecules¹⁸. Position clamp measurements on single kinesin motors gave stall forces that rose with ATP concentration, as anticipated (Fig. 3b). An inherent problem encountered in these experiments was the dramatic decline in kinesin processivity at loads approaching stall¹². Rapid detachment of beads from the microtubule made the collection of a large data set prohibitive, and accounted for much of the experimental scatter (Fig. 3b). To confirm the position clamp measurements, we also measured stall forces using a fixed optical trap. Kinesin-coated beads moving outwards in a stationary trap experience increasing loads until they arrive at a position where forward progress ceases: the final displacement, multiplied by the trap stiffness, supplies a measure of the stall force¹². Kinesin stall forces determined by this method exhibit a similar increase with ATP concentration (Fig. 3b).

How is the kinesin ATPase pathway coupled to mechanical movement under load? One possibility is that kinesin coupling is tight, involving no futile hydrolysis events that fail to generate movement, even under load (that is, $\epsilon = 1$). In this case, any changes in the rates of load-dependent transitions must account entirely for the observed decreases in velocity under load. However, more complicated scenarios may be entertained. Although the rise in K_m with load eliminates the simpler versions of loose coupling, mechanochemical coupling in kinesin might display an admixture of both tight and loose elements, with load-dependent transitions and unproductive hydrolyses conspiring jointly to lower the velocity. Another candidate mechanism for loose coupling involves an increase in the frequency of 8-nm backwards steps^{10–12} with load. However, the latter mechanism may be discounted because the fraction of backwards steps found in this study (data not shown) and by others remains only ~ 5 –10% at both negligible^{8,11} and substantial^{11,12} loads.

To study the coupling mechanism under load, we performed a fluctuation analysis^{23,24} of kinesin movement subjected to the force clamp. Fluctuation analysis, which relies on the processive property exhibited by some molecular motors, has been used previously to study ATP coupling in kinesin near zero load⁸, as well as the duty ratio of the bacterial rotary motor under torque²⁵. Fluctuations in displacement are quantified by a randomness parameter, r , which measures the temporal irregularity of mechanical advances. For an ensemble of displacement records, $x(t)$, r is defined by

$$r = \lim_{t \rightarrow \infty} \frac{\langle x^2(t) \rangle - \langle x(t) \rangle^2}{d \langle x(t) \rangle}$$

where d is the motor step size, and the angle brackets denote an ensemble average^{23,24}. A perfectly clocklike motor displays no stepping irregularity, and has $r = 0$, whereas a ‘Poisson’ motor with one biochemical transition moves at exponentially distributed intervals, and has $r = 1$. In general, r^{-1} provides a continuous measure of the number of rate-limiting transitions in the overall mechanochemical cycle. Determinations of randomness are insensitive to thermal or other stationary noise and can therefore be performed even when individual steps are not resolved. When steps can be discerned amid

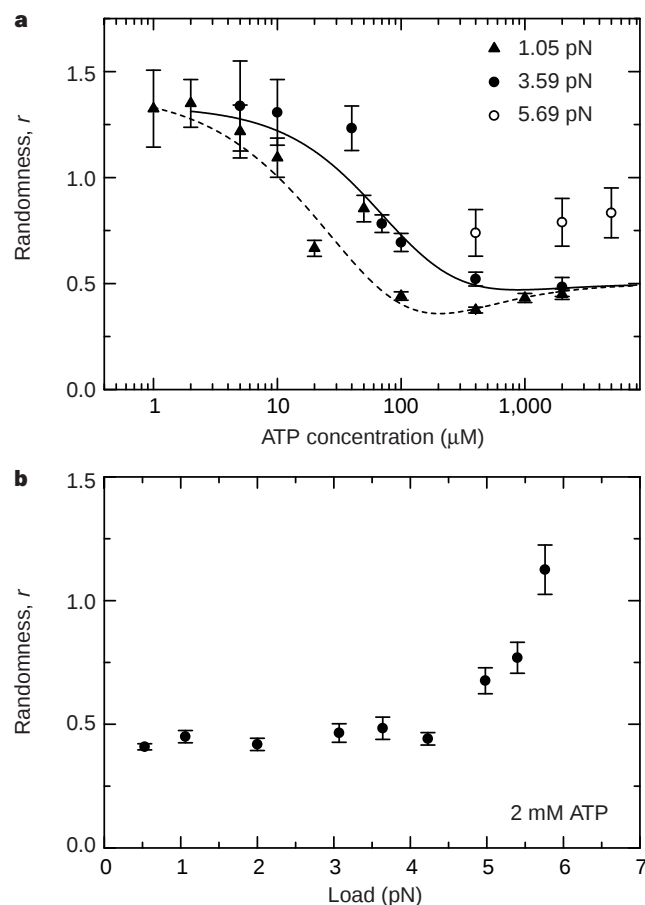


Figure 4 Displacement fluctuations. **a**, Semi-logarithmic plot of the randomness parameter, r (mean $\pm$ s.e.m.), versus ATP concentration for fixed loads (filled triangles, 1.05 pN $\pm$ 0.01, $N = 11$ –102; filled circles, 3.59 $\pm$ 0.03 pN, $N = 30$ –79; open circles, 5.69 $\pm$ 0.03 pN). The rise in r from $\sim\frac{1}{2}$ to ~ 1 begins at a higher ATP level for 3.59 pN than for 1.05 pN, reflecting the higher K_m at the former load. To guide the eye, data are compared to two-parameter analytical fits (dashed line, 1.05 pN; solid line, 3.59 pN). These fits were performed using the Michaelis-Menten parameters from Fig. 2, assuming two rate-limiting transitions at saturating ATP and one ATP hydrolysis per 8-nm step. Backward steps were assumed to occur in an ATP-dependent manner (see Methods). **b**, The randomness parameter, r (mean $\pm$ s.e.m.), versus load at 2 mM ATP (circles). Data points represent an average of 43–87 runs.

noise, fluctuation analysis is more robust than a direct analysis of stepping intervals, which suffers from the ‘missing event’ problem²³, an inability to distinguish the shortest steps. However, r is affected by phenomena that increase fluctuations at long time scales, such as backwards steps, enzyme inactivations, or futile hydrolyses. All of these can raise r , sometimes beyond unity^{8,23}.

The determination of r for negligible loads at limiting ATP verified that mechanochemical coupling was 1:1 in this regime⁸. Moreover, for minimal loads and saturating ATP, r was $\sim\frac{1}{2}$, which is indicative of two ATP-independent, rate-limiting transitions per step^{8,23}. But how is the randomness affected by the application of an external load? In the ‘tightly coupled’ scenario, the coupling is load-invariant by definition, so the ATP dependence of r must be correspondingly invariant, at least out to forces where there are still about two rate-limiting transitions. Tight coupling therefore predicts $r \approx \frac{1}{2}$ at saturating ATP, independent of load. In sharp contrast, ‘loosely coupled’ scenarios that invoke a decrease in coupling efficiency¹⁰ lead to a randomness that rises with load, in proportion to the concomitant decrease in velocity²³. For such mechanisms, values for r at saturating ATP should rise upwards

from $\frac{1}{2}$ towards 1 as load increases. Other ‘loosely coupled’ scenarios, such as those involving an increase in the frequency of backwards steps, should cause r to rise still more sharply. Coupling behaviour can therefore be studied by determining r as a function of load at saturating ATP levels. Nevertheless, at sufficiently high loads, even nominally tightly coupled schemes may lead to $r \approx 1$ at saturating ATP: this occurs for such high loads that the possibility for two rate-limiting transitions ultimately breaks down, as one or the other mechanically sensitive rate slows to become the sole determinant of velocity. At these extreme loads and beyond, one cannot distinguish between coupling mechanisms through measurements of randomness.

Measurements of the ATP dependence of randomness for low external load (1.05 $\pm$ 0.01 pN; Fig. 4) were similar to those obtained near zero load⁸: r rose from $\sim\frac{1}{2}$ at saturating ATP through 1 when ATP levels dropped below K_m (the increase in r above unity probably reflects the occurrence of 8-nm backwards steps, comprising 5–10% of all steps^{8,11,12}, as discussed⁸). A similar dependence of r upon ATP concentration was observed for a fixed load of 3.59 $\pm$ 0.03 pN (equal to about half the stall force), an observation that is consistent with tight coupling. Still stronger evidence in support of tight coupling came from the determination of randomness as a function of load for saturating ATP levels. The data of Fig. 4b show r to be nearly independent of load out to forces approaching ~ 5 pN: this invariance constitutes a clear signature of tight coupling, and can be used to place a theoretical bound on the fraction of total ATP hydrolyses that fail to produce movement. Even in conservative scenarios involving only 3% or 4% backwards steps, to remain consistent with these data the fraction of futile hydrolyses must be less than 15% or 10%, respectively, of all events. More realistic frequencies of backward stepping which are slightly higher^{8,11,12} lead to an even lower bound. (It is nevertheless possible to construct more complicated reaction schemes involving ~ 1 –2% backwards steps, plus more than four rate-limiting biochemical transitions under load, which admit a larger fraction of unproductive hydrolyses. However, we are inclined to exclude such complexities because they seem inconsistent with motility^{8,23,24} and solution^{21,22} studies suggesting that kinesin’s microtubule-stimulated ATPase has only about two rate-limiting transitions in the absence of load.) For forces approaching stall (beyond 5 pN; Fig. 4b), r rises steeply through 1, and these data alone do not indicate whether coupling remains tight. However, at 5.69 pN, we found no significant difference between r at 2 mM ATP and at an ATP level comparable to K_m (400 μM ; Fig. 4a), an observation consistent with a rise in futile hydrolyses at extreme loads. Because of the loss of processivity, however, we were unable to measure r for 5.69 pN at more than a few ATP levels (Fig. 4a). *In vivo*, detachment may serve to prevent kinesin motors from wasting ATP when faced with insurmountable loads².

The finding of tight coupling between ATP hydrolysis and mechanical stepping would seem to rule out many current theoretical models for force generation by kinesin. One class of models that is inconsistent with our findings predicts that most of the velocity decline under load is due to a loss of coupling efficiency and incorporates either pure loose coupling¹⁰, or at most one force-sensitive transition^{26–29}. Another apparently untenable class of models constitutes the so-called thermal ratchets (reviewed in ref. 30), for which the coupling efficiency depends sharply on load. However, we believe that the energy landscape picture proposed to explain the motion of RNA polymerase¹⁸ might be adapted to accommodate kinesin mechanochemistry.

Additional experimental work lends support to our conclusions. Recent studies of the average kinesin displacement under load following the release of caged ATP appear to be consistent with tight coupling³¹. Moreover, earlier experiments that tested the application of loads in the same direction as kinesin motion demonstrated an increase in velocity for both limiting and high

ATP levels¹². This counterintuitive finding may be explained by assuming that kinesin has two load-dependent transitions. The application of forward loads would raise the rates of both load-dependent transitions, leading to velocity increases when k_b is limiting at low ATP, but also when k_{cat} is limiting at high ATP.

Which transitions in the kinesin mechanical cycle are load-dependent? If load affects two (or more) biochemical transitions, one possibility is that each 8-nm kinesin mechanical step might comprise two shorter 'substeps' (ref. 32). However, although substeps represent transitions that seem likely to be load-dependent, they may lack the requisite ATP dependence, or they might be rapid and not rate-limiting even under load. Therefore, possible substeps in the 8-nm advance need not represent two load-dependent transitions. Conversely, a finding of two load-dependent transitions does not imply the existence of substeps: these transitions might correspond to biochemical events distinct from those which actually produce the forward progress, for example ATP binding and/or ADP release. Finally, because kinesin molecules carry two heads which are thought to operate 'hand over hand' in alternation, it remains an open question as to whether two load-dependent transitions take place within the same head or are instead shared between them. Further experiments will be required to determine the number and location of load-dependent transitions within the kinesin mechanochemical cycle. □

Methods

Assays. Motility assays were performed as described⁸. All chemicals were from Sigma, except glucose oxidase (250 $\mu\text{g ml}^{-1}$) and catalase (30 $\mu\text{g ml}^{-1}$), which were from Calbiochem.

Instrumentation and calibration. In brief, the optical trap was based on a Diaphot 200 inverted microscope (Nikon) outfitted to accommodate a 100 $\times$ /1.3 NA oil-immersion objective (Plan-Neofluar, Zeiss), into which a Nd:YVO₄ laser beam (Spectra Physics; 1,064 nm) was introduced through a custom laser port (Nikon). Trap position within the specimen plane was specified using two acousto-optical deflectors (IntraAction) digitally controlled by computer. Position detection was accomplished by focusing a low-power ($\sim 400 \mu\text{W}$ HeNe laser beam (Uniphase, 633 nm, fibre-coupled) onto an optically trapped 0.5 μm silica bead and measuring the deflected light in a plane conjugate to the back focal plane of the microscope condenser using a quadrant photodiode arrangement¹⁴. The trapping laser and detector beam foci were coaligned both axially and laterally, so that both video-based centroid tracking of a bead and photodiode-based detection of its position produced the same velocities and displacements as measured from the trap centre. This alignment was done using kinesin-coated beads attached to microtubules by 2 mM AMP-PNP, a non-hydrolysable ATP analogue. To simulate motility, the piezoelectric microscope stage was moved in either a triangle wave or a stochastic stepping pattern while the bead was held in the force clamp. The position detector was calibrated each time a bead was selected or the position in the field of view had changed¹⁴. Trap stiffness calibrations^{10,15} were done beforehand by determining the roll-off frequency of displacement power spectra for trapped beads as a function of laser power. To minimize differences in force arising from variation in the sizes of beads, we used a highly uniform population of silica spheres¹⁰ (gift from E. Matijevic). The standard deviation in trap stiffness (and therefore the force at a given displacement) from bead to bead was 4.2% ($n = 100$), based on measured power spectra obtained from 10 individual beads at each of 10 different laser light levels. Linearity of the trap force for displacements out to 200 nm from the trap centre was verified by applying a sinusoidal displacement to the microscope stage and recording the response of a trapped bead¹⁵.

To ensure accurate determinations of r , we performed control experiments analogous to those reported previously⁸. Kinesin-coated beads were placed on immobilized microtubules and motor activity was halted with AMP-PNP. To simulate single kinesin motion, the piezoelectric microscope stage was moved in discrete steps of 7–8 nm, at either fast ($\sim 660 \text{ nm s}^{-1}$) or slow ($\sim 14 \text{ nm s}^{-1}$) average speeds, at either high ($\sim 6 \text{ pN}$) or low ($\sim 1 \text{ pN}$) constant loads, and with stepping intervals chosen either from an exponential distribution ($r = 1$) or from a convolution of two exponentials ($r = \frac{1}{2}$). Under all eight possible conditions, control determinations of randomness and velocity were at worst

within $\sim 15\%$ and $\sim 5\%$, respectively, of their nominal values (data not shown). Before control and experimental runs, each bead was held $\sim 300 \text{ nm}$ above the microtubule and the position detector response was calibrated on-line^{5,14}. To randomize measurement errors arising from vertical offsets, the microscope was refocused between runs such that the microtubule image was sharp when viewed by video-enhanced microscopy.

Records of bead displacements along microtubules that were aligned with a single axis (x) of the position detector were sampled at 19–21 kHz, anti-alias filtered at 10 kHz, digitized at 12-bit resolution, used for trap feedback, then decimated and saved to disk at 1.9–2.1 kHz. The force clamp thus operates along the x -axis only. The position of the optical trap was updated at 19–21 kHz but saved at 190–210 Hz for better computer performance. The running bead position was integrated with a variable (3–50 ms) boxcar window that was used to compute the updated trap position. We optimized the integration window during all control and experimental runs to improve feedback stability. Because the force clamp responds to variations in bead position averaged over the integration time, high-frequency force variations are permitted. To minimize variations caused by brownian motion of the bead, we set the bead-trap separation at 175 nm for all loads, except for the lowest loads of $\sim \frac{1}{2} \text{ pN}$ and $\sim 1 \text{ pN}$, for which the separation was 100 nm to maintain a minimum trap stiffness. The load for each run was computed off-line from the known stiffness and the difference in the digitized bead and trap positions (Fig. 1). All runs used for analysis were required to lie laterally within $\pm 50 \text{ nm}$ of the microtubule axis. Upon activation of the feedback control program, a kinesin-coated bead was pulled to one edge of the detection zone, allowed to bind to the microtubule, and then to proceed towards the opposite edge of the zone: the load remained clamped over the interval (-150 nm , $+150 \text{ nm}$) from the zone centre.

Stall force measurements using the position clamp were calibrated and performed as described¹⁷. In brief, position measurements were made with an optical trapping interferometer³ distinct from the instrument used for the rest of this work, taking into account the slight variations in equilibrium bead position with laser power (these effects were calibrated before stall measurements were taken)¹⁷. Kinesin-driven beads were stalled at $\sim 105 \text{ nm}$ from the trap centre, and stalls were required to last $> 2 \text{ s}$ to be scored. By multiplying this distance from the trap centre, the measured laser power at stall, and the predetermined trap stiffness ($4.5 \times 10^{-4} \text{ pN nm}^{-1} \text{ mW}^{-1}$), stall forces were calculated. Stall force measurements using a fixed optical trap were done with the force-clamp instrument in its open-loop mode. Again, stalls were required to last for $> 2 \text{ s}$ to be scored.

Analysis. Mean velocities, variance and randomness were determined using custom software written in LabView 5.0 (National Instruments) as described^{8,23}, except for the following. Experimental and control displacement records within $\pm 150 \text{ nm}$ of the detector centre were used. For inclusion in velocity analysis, runs had to last at least 100 nm ($> 5 \text{ pN}$) or 300 nm ($< 5 \text{ pN}$). Only runs over the full 300-nm range were used for the determination of r , except for the three r values at $5.69 \pm 0.03 \text{ pN}$ (Fig. 4a), for which we used runs of 150 nm minimum, because of the reduced processivity at high load. To determine the linear slope of the displacement variance versus time, a linefit was made to the variance over the first 40 nm/ $\langle v \rangle$ time interval, excluding at least the first $\sim 15 \text{ ms}$ (the correlation time of the bead position in the force clamp). Experimental and control runs at limiting ATP levels were smoothed and resampled with a 10–25 ms window, a procedure that improves computational speed but does not affect r (ref. 8). To guide the eye, analytical fits to r versus ATP were made under the assumption of two rate-limiting transitions at saturating ATP and one ATP hydrolysis per 8-nm step⁸. To account for backwards steps, we fit to the formula⁸ $r = (\bar{r} - 1)(1 - 2P_-) + 1/1 - 2P_-$, where $\bar{r}$ is the randomness in the absence of backwards steps, and P_- , the fraction of steps backwards, is inversely proportional to a linear function of [ATP].

Received 16 March; accepted 10 May 1999.

- Howard, J. Molecular motors: structural adaptations to cellular functions. *Nature* **389**, 561–567 (1997).
- Howard, J. The movement of kinesin along microtubules. *Annu. Rev. Physiol.* **58**, 703–729 (1996).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Howard, J., Hudspeth, A. J. & Vale, R. D. Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–158 (1989).

5. Visscher, K. & Block, S. M. Versatile optical traps with feedback control. *Methods Enzymol.* **298**, 460–489 (1998).
6. Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
7. Ishijima, A. *et al.* Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* **92**, 161–171 (1998).
8. Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
9. Hua, W., Young, E. C., Fleming, M. L. & Gelles, J. Coupling of kinesin steps to ATP hydrolysis. *Nature* **388**, 390–393 (1997).
10. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).
11. Kojima, H., Muto, E., Higuchi, H. & Yanagida, T. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* **73**, 2012–2022 (1997).
12. Coppin, C. M., Pierce, D. W., Hsu, L. & Vale, R. D. The load dependence of kinesin's mechanical cycle. *Proc. Natl Acad. Sci. USA* **94**, 8539–8544 (1997).
13. Meyhöfer, E. & Howard, J. The force generated by a single kinesin molecule against an elastic load. *Proc. Natl Acad. Sci. USA* **92**, 574–578 (1995).
14. Visscher, K., Gross, S. P. & Block, S. M. Construction of multiple-beam optical traps with nanometer-resolution position sensing. *IEEE J. Sel. Topics Quantum. Electron.* **2**, 1066–1076 (1996).
15. Svoboda, K. & Block, S. M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994).
16. Finer, J. T., Simmons, R. M. & Spudich, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
17. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1996).
18. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
19. Howard, J. The mechanics of force generation by kinesin. *Biophys. J.* **68**, 245s–255s (1995).
20. Hackney, D. Kinesin ATPase: Rate-limiting ADP release. *Proc. Natl Acad. Sci. USA* **85**, 6314–6318 (1988).
21. Gilbert, S. P., Webb, M. R., Brune, M. & Johnson, K. A. Pathway of processive ATP hydrolysis by kinesin. *Nature* **373**, 671–676 (1995).
22. Ma, Y. Z. & Taylor, E. W. Mechanism of microtubule kinesin ATPase. *Biochemistry* **34**, 13242–13251 (1995).
23. Svoboda, K., Mitra, P. P. & Block, S. M. Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl Acad. Sci. USA* **91**, 11782–11786 (1994).
24. Schnitzer, M. J. & Block, S. M. Statistical kinetics of processive enzymes. *Cold Spring Harbor Symp. Quant. Biol.* **60**, 793–802 (1995).
25. Samuel, A. D. T. & Berg, H. C. Torque-generating units of the bacterial flagellar motor step independently. *Biophys. J.* **71**, 918–923 (1996).
26. Leibler, S. & Huse, D. A. Porters versus rowers: a unified stochastic model of motor proteins. *J. Cell Biol.* **121**, 1357–1368 (1993).
27. Peskin, C. & Oster, G. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophys. J.* **68**, 202s–211s (1995).
28. Duke, T. & Leibler, S. Motor protein mechanics: a stochastic model with minimal mechanochemical coupling. *Biophys. J.* **71**, 1235–1247 (1996).
29. Derényi, I. & Vicsek, T. The kinesin walk: a dynamic model with elastically coupled heads. *Proc. Natl Acad. Sci. USA* **93**, 6775–6779 (1996).
30. Astumian, R. D. Thermodynamics and kinetics of a brownian motor. *Science* **276**, 917–922 (1997).
31. Higuchi, H. & Yanagida, T. Force generation and detachment of single kinesin molecules activated by laser photolysis of caged ATP and ADP. *Cell Struct. Funct.* **23**, suppl., 198 (1998).
32. Coppin, C. M., Finer, J. T., Spudich, J. A. & Vale, R. D. Detection of sub-8-nm movements of kinesin by high-resolution optical-trap microscopy. *Proc. Natl Acad. Sci. USA* **93**, 1913–1917 (1996).

Acknowledgements. We thank Y. Inoue, M. Nishiyama, L. Satterwhite and M. Wang for discussions; J. de Georgis for squid dissection; and D. Peoples for machining. This work was supported by grants to S.M.B. from the NIGMS, NSF and W. M. Keck Foundation, predoctoral fellowships to M.J.S. to the American Heart Association, the Charlotte Elizabeth Proctor Fund, and the Program in Mathematics and Molecular Biology Burroughs Wellcome Fund, and a postdoctoral fellowship to K.V. from the Burroughs Wellcome Fund of the Life Sciences Research Foundation.

Correspondence and requests for materials should be addressed to M.J.S. (e-mail: schnitzer@physics.bell-labs.com).

Cells, gels and wells

Warming to the theme of biotechnology, a new heatsealing device handles PCR plates, and DNA is separated, amplified, purified and sequenced.

Cloning disks

From Bel-Art

www.belart.com

Sterile cloning disks, to save time when isolating large numbers of clones

Scienceware sterile paper cloning disks are an alternative to cloning cylinders for isolating cell colonies for transfer to 24-well plates. Using a procedure developed by Domann and Martinez (www.belart.com/catalog/lifescience/206-sterile-clone.disc.html) reduces the risk of contamination and drying, because cells are exposed for a much shorter period of time than with cloning cylinders. The disks are available in three sizes, and are RNase and DNase free.

Reader Service No. 100

ssDNA

From Labsystems

www.labsystems.fi

Easy purification of PCR-amplified biotinylated DNA

New from Labsystems is a range of single-stranded DNA reagents based on streptavidin-coated magnetic particles for easier purification of PCR-amplified biotinylated DNA. The reagents are optimized for use in Labsystems' KingFisher automated processor. The templates can be washed free from the PCR components and rendered single stranded without the need for extraction, washing and centrifugation. The pure ssDNA produced can then be used for further studies, such as DNA sequencing.

Reader Service No. 101

Presto workstation

From Zymark

www.zymark.com

Automation for 96- and 384-well plates

The Presto Liquid Handling Workstation is a designed for microplate applications, minimizing the errors and workload involved in pipetting into high-density plates. Two pipette arms and individually controlled syringe pumps assure fast, precise liquid transfers. Software features include a graphical user interface and point-and-click controls. The device has an ultra-small foot print so can easily be used on a standard lab bench-top; it is available as a stand-alone workstation or on Integrated Assay Systems.

Reader Service No. 102

ALPS

From Abgene

Plate sealing, suitable for automated systems

The ALPS, Automated Laboratory Plate



Clone arranger – these look good on paper

Sealer, performs rapid heatsealing of deep-well, PCR and microtitration plates. The design accommodates plates of various heights and sizes. The unsealed plate is placed on a shuttle, which is drawn into the unit automatically and the plate is sealed in 10–15 seconds. The shuttle then returns to its starting position, allowing the plate to be removed by a robotic arm. One roll of sealing film can seal approximately 4,000 plates.

Reader Service No. 103

Transgenic proteins

From Genzyme/Invitrogen

www.genzyme.com

A milk expression vector kit for transgenic protein production

Genzyme Transgenics Corporation and Invitrogen Corporation earlier this year launched the pBC1 Milk Expression Vector Kit, providing proprietary technology for the production of recombinant proteins in milk of transgenic mice. The kit contains Genzyme Transgenics' β -casein expression promoter and other DNA sequences to enable researchers to insert a gene into the DNA of a fertilized mouse ovum. Use of the pBC1 kit is limited to the generation of transgenic mice for early-stage feasibility studies. If initial feasibility studies are favourable, researchers would be required to contract with Genzyme Transgenics for commercial use.

Reader Service No. 103A



Liquid launch: Presto stands alone

Latitude

From FMC BioProducts

www.fmc.com

Run 40 lanes in 30 minutes with Latitude precast midgels.

Latitude Precast Agarose Midgels provide separation of DNA ranging from 10 bp to greater than 10 kb. The gels come ready for use 20- and 40-well formats. For maximum throughput, these gels are compatible with alternate-well microtitre plate spacing. Latitude gels are available with a variety of agarose concentrations and buffers.

Reader Service No. 104

Automated sequencing

From Sigma

www.sigma.sial.com

A new brochure on automated DNA sequencing

Sigma's new guide to automated sequencing products includes details of the company's high-throughput acrylamide solutions, AutoPAGE 4.25% and AutoPAGE Plus 4.5%. These solutions and their optimized protocols are claimed to yield more data in less time than standard gel runs. Other Sigma products covered include Premixed 10× TBE qualified for automated sequencing, and halfTerm and halfBD solutions to reduce terminator premix costs by at least 50%.

Reader Service No. 105

ChemScanRDI

From Chemunex

A new approach to testing for microbial contaminants in mammalian cell production

ChemScanRDI is a solid-phase laser scanning



Time to gel

cytometer that can directly detect a single bacteria among 10^6 mammalian cells within 90 minutes of sampling. Chemunex say that more than 40 pharmaceutical companies are using the system to detect micro-organisms. ChemScanRDI avoids the need for a cell multiplication step, allowing early assessment of any microbial contamination. **Reader Service No. 106**

Biotin-RNA

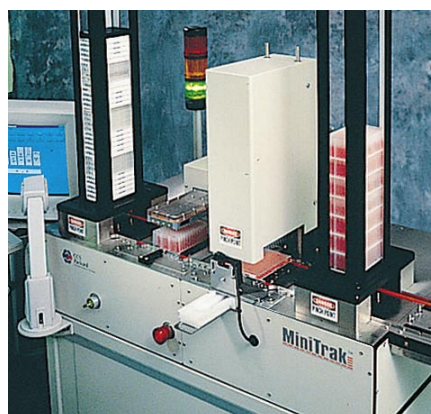
From Enzo Diagnostics, Inc. www.enzobio.com
High-yield production of biotin-labelled RNA

The BioArray HighYield system is an RNA transcript labelling system for the production of large amounts of biotin-labelled RNA for use in nucleic acid array analyses. The BioArray HighYield kit makes use of T7 RNA polymerase for the production of up to 100 µg or more of biotin-labelled RNA per reaction. The HighYield system includes all labelling reagents and the labelled nucleotides Bio-UTP and Bio-CTP. **Reader Service No. 107**

MiniTrak

From CCS Packard www.packardinst.com
Microplate processing

MiniTrak is a compact four-module microplate processing system with two microplate stacker cassettes, a 96-tip pipette



Mini-Trak: four modules for flexibility

dispenser and an open module for future expansion. MiniTrak is compatible with 96- and 384-well plates, and other formats including deepwell microplates. The open module provides the flexibility to run an optional 96- or 384-tip dispenser, wash module or other specialized attachment. **Reader Service No. 108**

DuoLuX

From Vector www.netwiz.net/ vector
High-power chemiluminescence

The DuoLuX Chemiluminescent Substrate is claimed to provide better light emission, higher sensitivity and greater flexibility than most other chemiluminescent substrates. **Reader Service No. 109**

Based on acridan chemistry, the same DuoLuX Substrate can be used with either alkaline phosphatase or peroxidase enzyme systems in Southern, northern, western and dot blots, or in ELISA applications. Film can be exposed to blots up to 8 hours after substrate development, offering convenience and the flexibility to optimize band intensities or resolution. **Reader Service No. 109**

CDP-Star

From Roche Diagnostics www.roche.com
Chemiluminescence, ready to go

This ready-to-use chemiluminescent substrate for alkaline phosphatase offers reproducible detection of non-radioactively labelled biomolecules in solution or on solid supports. 'CDP-Star, ready-to-use' is recommended for a range of gene analysis applications, including northern, western and Southern blot techniques. The new dioxetane substrate is supplied in 50-ml dropper bottles and is stable for 18 months at 4 °C. **Reader Service No. 109a**

T-REx

From Invitrogen www.invitrogen.com
A high-level regulated mammalian expression system

Invitrogen say that their new T-REx (tetra-cycline-regulated expression) System is the first inducible mammalian expression system to use the full-length CMV promoter. The CMV promoter provides high yields of recombinant protein. T-REx uses a repressor to regulate transcription, eliminating non-specific effects of transactivators. **Reader Service No. 110**

CyclePlate-384DW

From Robbins Scientific www.robsci.com
PCR plates for high-throughput genomics

CyclePlate-384DW (discrete well) Thin Wall Plates are built to retain industry-standard dimensions before and after thermal cycling. CyclePlate-384DW plates are sealed with CycleSeal plate sealer and are compatible with the 384-well heating blocks used on

ADVERTISEMENTS

Diagnostic Reagents

Pefachrome®-Series: chromogenic and fluorigenic substrates

Pefabloc®-Series: serine proteinase inhibitors

Purified snake venom proteins (e.g. Protac®, Ecarin, Batroxobin)

PENTAPHARM LTD

CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

Customised monoclonal & polyclonal antibodies with guarantee!

BioGenes GmbH
Köpenicker Str. 325 · D-12555 Berlin
Tel. +49-30-65 76 23 96 Fax: - 97
E-mail: BioGenes@Berlin.sireco.net
Internet: <http://www.biogenes.de>

READER ENQUIRY NO. 17

**THIOREDOXIN
GLUTAREDOXIN
(E.coli,mammalian)**

WWW.imcocorp.se

IMCO Corporation Ltd AB

Fax: +46 8 728 47 76 - email: info@imcocorp.se

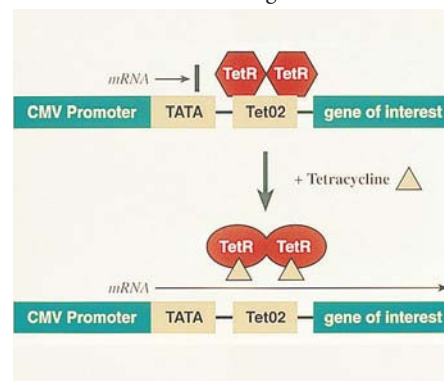
READER ENQUIRY NO. 18

**Lysyl Endopeptidase
(Achromobacter Protease 1)**
Specifically cleaves all Lys-X bonds including Lys-Pro.
Highly Stable, Water Soluble

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 19



Express yourself, using T-REx

both the MJ Research and Perkin-Elmer thermal cyclers.

Reader Service No. 111

MACS mRNA isolation

From Miltenyi Biotec www.miltenyibiotec.com

mRNA from total RNA in 8 minutes

The MACS mRNA Isolation Kit uses MACS magnetic separation technology for large-scale isolation of mRNA from cells, animal or plant tissue. The kit contains reagents, Oligo (dT) MicroBeads, large-scale LysateClear columns and MACS M columns sufficient for 8 separations, each for up to 50 µg mRNA from a maximum of 5×10^7 cells, 250 mg animal tissue or 500 mg plant tissue. The isolated mRNA is rRNA free and suitable for a wide range of downstream applications.

Reader Service No. 112

BESS-T&G

From Epicentre www.epicentre.com

Mutated base identification and DNA typing

BESS-T&G Base Reader Kits provide T- and G-lane sequence from PCR products made using labelled primers, without performing dideoxy sequencing. If the PCR product is made using two different dye-labelled primers, the T or G sequence of both DNA template strands can be read simultaneously. For mutation detection, Base Excision Sequence Scanning (BESS) scans can provide confirmation of dideoxy sequencing results.

Reader Service No. 113

µPAGE-10

From J&W Scientific, Inc. www.jandw.com

Kits for making sense of antisense

Designed for high-resolution separations of antisense therapeutic agents, primers and probes, the µPAGE-10 gel-filled capillary kit for capillary electrophoresis contains everything required for performing an analysis: three CE columns, buffer, and a test mixture.

Spare columns and buffers can be ordered.

Reader Service No. 114

ScreenReady Targets

From BioSignal, Inc. www.biosignal.com

Targets for high-throughput screening

Montreal-based company BioSignal Inc. has introduced the first generation of its ScreenReady Targets product line for high throughput screening (HTS). ScreenReady Targets consist of SignalScreen G-protein-coupled receptor membranes (GPCRs) pre-coated onto basic 96-well FlashPlate microplates and supplied with the appropriate NEN radioligand and assay protocol. ScreenReady Targets are fully automatable and stable for 1 month at 4 °C,

Reader Service No. 115

PCR amplification

From Life Technologies www.lifetech.com

A new PCR amplification enzyme

The Platinum Pfx DNA polymerase, introduced in May this year, combines high yield with specificity in PCR procedures. The enzyme preparation includes an automatic 'hot start' capability to reduce misprimings and nonspecific amplification. It will amplify genomic templates up to 12 kb and plasmid templates up to 20 kb, and is recommended for site-directed mutagenesis.

Reader Service No. 115A



Ready-reckoner: Biosignal, on targets

Nytran SuPerCharge

From Schleicher & Schuell www.s-and-s.com

A nylon membrane with high charge density and low background binding

Nytran SuPerCharge combines a higher binding capacity than typical nylon membranes with lower background binding, to give good signal-to-noise ratios and high sensitivity. In addition, the clean surface and uniform morphology of these membranes are claimed to give excellent colony growth. Nytran SuPerCharge and Nytran N membranes (a lower-charge version) are recommended for Southern, northern, dot/slot-blots and colony/plaque lifts. They are compatible with alkaline transfer conditions and UV crosslinking, and can be used with isotopic or non-isotopic detection.

Reader Service No. 115B

Poster session

From Pierce Chemical www.piercenet.com

A poster and handbook for those working with substrates

Two new reference tools are available free from Pierce Chemical Company: a substrates selection guide poster and a Western Blotting Handbook. The poster includes details of the features of Pierce soluble and precipitating substrates, based on the enzyme that each detects or the type of signal generated (chemiluminescent, colorimetric or chemifluorescent.) The 24-page Western Blotting Handbook includes abbreviated and full-length Western blotting protocols, as well as recommendations for antibody concentration and blocking buffer optimization.

Reader Service No. 116

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

contact : lightcycler@TIB-MOLBIOL.de

TIB
MOLBIOL

HybProbes

for the **LightCycler**

- Mutation Detection
- Quantitative PCR

Custom Synthesis of Hybridization Probes for the new LightCycler under licence of Roche Diagnostics

READER ENQUIRY NO. 7

Check out our homepage at <http://www.resgen.com>

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 30

© 1999 Macmillan Magazines Ltd

Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check Us Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com

New Probe Sizes
0.45cm

READER ENQUIRY NO. 54